ELEC/STAT 321

Stochastic Signals and Systems

Charles Clayton

February 27, 2017

Table of Contents

List of Figures

List of Tables

List of Equations

Chapter 1

Intro From Statistics Department

1.1 Probability

A **random experiment** is when an outcome cannot be determined *exactly* beforehand. The **sample space**, Ω, is all the possible outcomes of the random experiment. So you might not know whether or not the alcohol content of a beer your brewing will be 5% or 6%, but you know $\Omega = [0, 100\%]$.

Subsets of Ω are called **events**, denoted as A, B, C, etc. An event **occurs** if the outcome, ω, is in the event. That is, $\omega \in A$. For instance, in our beer-brewing example, the event could be that $A = [3, 9\%]$. If the outcome is 5%, then A occurs.

A **sigma-field**, $\mathcal{F}$, is a collection of events that we can assign probabilities to. At minimum, the sample space and the empty-set must be elements of the sigma-field.

$$\emptyset, \Omega \in \mathcal{F}$$

If the event A is in the sigma field, then $\overline{A}$ must also be in the sigma field.

$$A \in \mathcal{F} \implies \overline{A} \in \mathcal{F}$$

The sigma-field is the domain of the **probability function**, P, because P transforms these collections of events into probabilities.

$$P : \mathcal{F} \to [0, 1]$$

Note that the probability function is not like other functions, the input is not a number it is a set of events. The probability of the entire sample space is 1.

$$P(\Omega) = 1$$

For disjoint events A_n, the probability of the union of the events will be the sum of the probabilities of each event.

$$A_n \text{ disjoint} \implies P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$$

1.1.1 Some results to use

$$\boxed{P(\overline{A}) = 1 - P(A)} \tag{1.1}$$

Probability of complement

$$\boxed{A \subset B \implies P(A) \leq P(B)} \tag{1.2}$$

Probability of subset

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)} \tag{1.3}$$

Probability of union

Intuitively, (3) makes sense because if you have two Venn diagrams that overlap somewhat, to determine the area of the entire area, you add the two circles A and B together, then subtract the overlapping area $(A \cap B)$ because when you add A and B you count that area twice.

$$\boxed{P(\bigcup_{n=1}^{n} A_n) \leq \sum_{n=1}^{n} P(A_n)} \tag{1.4}$$

Boole's inequality

Where the events A_n are disjoint, then the inequality in equation 1.4 becomes an equality.

$$\boxed{P(A) = \frac{\#\text{ of outcomes in } A}{\#\text{ of outcomes in } \Omega}} \tag{1.5}$$

Probability of equally likely outcomes

The number of ways you can choose k items from a set of n items.

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!}} \tag{1.6}$$

Combinatorial formula

Example

Consider the 6/49 lottery. There is a box with 50 balls numbers 0-49, and 6 balls are chosen at random. What are the odds that of the 6 numbers you choose, x of those numbers are the picked balls.

—

The numerator is the number of possible combinations of draws that give you exactly x matches. If you are matching x numbers, then you are not matching $6 - x$ numbers. So of 6 you will have to choose x, and of the remaining 44 you will have chosen $6 - x$.

The denominator is then the number of all possible ways you can draw numbers.

$$p(x) = \frac{\binom{6}{x}\binom{44}{6-x}}{\binom{50}{6}}$$

So the probability of getting all 6 correct is 1/15,890,700 = 0.0000000063

1.2 Conditional Probability

The idea here is that although the outcome can be any element in the sample space, Ω, the range of possible outcomes can be narrowed down with the help of partial information, called a **conditioning event**. For instance, if we know that the beer we brewed is a weak beer, this is a conditioning event and we can narrow down the outcome from $[1 - 4\%]$.

Where $P(A)$ is the event of interest and B is the partial information.

$$\boxed{P(A|B) = P(A) \text{ given } B = \frac{P(A \cap B)}{P(B)}}$$

$$\boxed{P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B)}$$

$$\boxed{P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)}$$

$$\boxed{P(A|B) = 1 - P(\overline{A}|B)} \tag{1.7}$$

Conditional probabilities

Our previous results still hold with conditional probabilities:

$$P(\Omega|B) = 1$$

$$P(A|B) \geq 0$$

If A_n are disjoint, then

$$P\left(\bigcup A_n|B\right) = \sum P(A_n|B)$$

Example

Without any information, the probability that you get an A in your class is the probability that your grade, $g = [80, 100]$ out of possibilities $[0, 100]$.

$$P(\text{Getting an A}) = \frac{P([80,100])}{P([0,100])} = \frac{2}{10} = 0.2$$

However with the partial information that you will get at least a B, so $g \geq 70$, we can use that as our conditioning event.

$$= \frac{P(A \cap B)}{P(B)} = \frac{P([80,100] \cap [70,100])}{P([70,100])} = \frac{P([80,100])}{P([70,100])} = \frac{2}{3} = 0.667$$

1.2.1 Screening Tests

Conditional probability can be used for screening tests. If the test is positive, the condition is present. Consider testing manufactured items for defects. The item can be defective, D, or non-defective $\overline{D}$.

If an item is defective, the probability that the test will sense that (test positive) is the **sensitivity** of the test. This is how often your test will sense the defect.

$$\boxed{P(B|D)} \tag{1.8}$$

Sensitivity

If an item is non-defective, the probability that the test will be negative is called the **specificity**.

$$\boxed{P(\overline{B}|\overline{D})} \tag{1.9}$$

Specificity

Example

Say the sensitivity of your test is 0.95, the specificity is 0.99, and the probability of products actually being defective is $P(D) = 0.02$. What is the probability that the test will be incorrect?

—

Where D indicates the item is defective, and B indicates the test is positive:

$$P(B|D) = 0.95,\ P(\overline{B}|D) = 1 - 0.95 = 0.05$$

$$P(\overline{B}|\overline{D}) = 0.99,\ P(B|\overline{D}) = 1 - 0.99 = 0.01$$

The probability of the test testing positive:

$$P(B) = P(B \cap D) + P(B \cap \overline{D}) = P(D)P(B|D) + P(\overline{D})P(B|\overline{D})$$

$$= (0.02)0.95 + (1 - 0.02)0.01 = 0.0288$$

Probability of item being defective if the test is positive (accurate positive):

$$P(D|B) = \frac{P(B \cap D)}{P(B)} = \frac{(0.02)0.95}{0.0288} = 0.6597$$

Probability of item being defective if the test is negative (accurate negative):

$$P(D|\overline{B}) = \frac{P(\overline{B} \cap D)}{P(\overline{B})} = \frac{P(D)P(\overline{B}|D)}{P(\overline{B})} = \frac{(0.02)0.05}{1 - 0.0288} = 0.00103$$

The probability of the test being incorrect (false positive + false negative):

$$P(B \cap \overline{D}) + P(\overline{B} \cap D) = P(\overline{D})P(B|\overline{D}) + P(D)P(\overline{B}|D) = (0.98)0.01 + (0.02)0.05 = \mathbf{0.0108}$$

—

The screening test formula can summarized with the Bayes' Formula.

$$\boxed{P(D|B) = \frac{P(D \cap B)}{P(B)} = \frac{P(B|D)P(D)}{P(B|D)P(D) + P(B|\overline{D})P(\overline{D})}} \tag{1.10}$$

Simple Form of Bayes' Formula

Generalized, the formula looks like so:

$$\boxed{P(D_i|B) = \frac{P(D_i \cap B)}{P(B)} = \frac{P(B|D_i)P(D_i)}{\sum_{j=1}^{k} P(B|D_j)P(D_j)}} \tag{1.11}$$

General Form of Bayes' Formula

Where $D_1,\ D_2\ \ldots\ D_k$ is a **partition** of sample space Ω, meaning each D_i is disjoint ($D_i \cap D_j = \emptyset$ for $i \neq j$) and each outcome in the sample space is accounted for ($D_1 \cup D_2 \cup \cdots \cup D_k = \Omega$).

Let's use this iterative formula with another example.

Example

The items to be tested have two components, c_1 and c_2, say an IC wafer and a package. Suppose

$D_1 = \{\text{Only component } c_1 \text{ is defective}\}\ ,\ P(D_1) = 0.01$

$D_2 = \{\text{Only component } c_2 \text{ is defective}\}\ ,\ P(D_2) = 0.008$

$D_3 = \{\text{Both components are defective}\}\ ,\ P(D_3) = 0.002$

$D_4 = \{\text{Both components are non-defective}\}\ ,\ P(D_4) = 0.98$

$B = \{\text{Screening test is positive}\}$

And $P(B|D_1) = 0.95,\ P(B|D_2) = 0.96,\ P(B|D_3) = 0.99,\ P(B|D_4) = 0.01$

What is the probability of testing positive? What is the probability component c_1 is defective when the test is positive? c_2? What is the probability that the item is defective when the test results negative? What is the probability both components are defective when the test results positive? What's the probability of a testing error?

—

Probability of testing positive:

$$P(B) = \sum_{i=1}^{4} P(B|D_i)P(D_i) = 0.95(0.01)+0.96(0.008)+0.99(0.002)+0.01(0.98) = \mathbf{0.02896}$$

Note that the probability of defects is:

$$P(D) = P(D_1) + P(D_2) + P(D_3) = 0.01 + 0.008 + 0.002 = 0.02$$

So our test will have false positives.

Probability of defect in only c_1 is:

$$P(D_1|B) = \frac{P(D_1 \cap B)}{P(B)} = \frac{P(B|D_1)P(D_1)}{P(B)} = \frac{0.95(0.01)}{0.02896} = \mathbf{0.32804}$$

Probability of defect in only c_2:

$$P(D_2|B) = \frac{P(D_2 \cap B)}{P(B)} = \frac{P(B|D_2)P(D_2)}{P(B)} = \frac{0.96(0.008)}{0.02896} = \mathbf{0.26519}$$

Probability of defect in both c_1 and c_2:

$$P(D_3|B) = \frac{P(D_3 \cap B)}{P(B)} = \frac{P(B|D_3)P(D_3)}{P(B)} = \frac{0.99(0.002)}{0.02896} = \mathbf{0.06837}$$

Probability neither c_1 or c_2 are defective when the test is positive:

$$P(D_4|B) = \frac{P(D_4 \cap B)}{P(B)} = \frac{P(B|D_4)P(D_4)}{P(B)} = \frac{0.01(0.98)}{0.02896} = \mathbf{0.33840}$$

Redo substituting B with $\overline{B}$ to determine the probabilities when the test is negative.

The probability of error is then:

$$P(D_4|B) + P(D_1|\overline{B}) + P(D_2|\overline{B}) + P(D_3|\overline{B})$$

1.2.2 Independence

When A and B are independent events, the probability of A intersect B is the product of their probabilities:

$$\boxed{A \text{ and } B \text{ are independent} \iff P(A \cap B) = P(A)P(B)} \tag{1.12}$$

Condition for independence

If A and B are independent, then:

$$\boxed{P(A|B) = P(A) \text{ and } P(B|A) = P(B)} \tag{1.13}$$

Conditional probabilities with independence

Note that if events A and B are disjoint (mutually exclusive) or A is a true subset of B then they are dependent:

$$A \cap B = \emptyset \implies A \text{ and } B \text{ are not independent}$$

$$A \subset B \implies A \text{ and } B \text{ are not independent}$$

So in order to be independent, A and B must have some overlap, but one must not be completely within the other.

Example

Suppose $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, and $A =$1,2,3,4,5 and $B =$2,4,6,8.

If p is equally likely for each number, are A and B independent? What if $p(i) = \frac{i}{55}$?

—

If all numbers are equally likely, then:

$$P(A \cap B) = P(\{2, 4\}) = \frac{2}{10} = 0.2$$

Now note that:

$$P(A)P(B) = \frac{5}{10} \times \frac{4}{10} = 0.2$$

And so A and B are independent.

If $P(i) = \frac{i}{55}$, then:

$$P(A \cap B) = P(\{2, 4\}) = \frac{2 + 4}{55} = 0.10909$$

Whereas:

$$P(A)P(B) = \frac{1 + 2 + 3 + 4 + 5}{55} \times \frac{2 + 4 + 6 + 8}{55} = \frac{15}{55} \times \frac{20}{55} = 0.99174$$

And so A and B are not independent.

—

For more than three events, say events A_1, A_2, ... A_n, the condition for independence is:

For all $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ and all $1 \leq k \leq n$

$$\boxed{P(A_{i_1}) \cap P(A_{i_2}) \cap \cdots \cap P(A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k})} \tag{1.14}$$

General condition for independence

Meaning that if $n = 3$, the following by itself is insufficient:

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

The following are also required conditions:

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$

1.2.3 Reliability

Example

Suppose we have three components a, b, c, where A, B, C are the probability that each component works. We assume A, B, and C are independent.

Where $P(A) = P(B) = P(C) = 0.95$, calculate the reliability of the system:

$$\longrightarrow \boxed{\mathsf{a}} \longrightarrow \boxed{\mathsf{b}} \longrightarrow \boxed{\mathsf{c}} \longrightarrow$$

$$P(\text{System works}) = P(A \cap B \cap C) = P(A)P(B)P(C) = 0.95^3 = 0.857$$

Now calculate the reliability of the system:

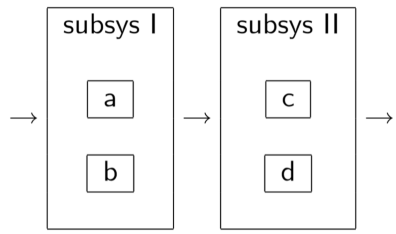
$$\longrightarrow \boxed{\begin{array}{c} \boxed{\mathsf{a}} \\ \\ \boxed{\mathsf{b}} \\ \\ \boxed{\mathsf{c}} \end{array}} \longrightarrow$$

$$\begin{aligned} P(\text{System works}) &= 1 - P(\text{System fails}) \\ &= 1 - P(\overline{A} \cap \overline{B} \cap \overline{C}) \\ &= 1 - P(\overline{A})P(\overline{B})P(\overline{C}) \\ &= 1 - (1 - 0.95)^3 \\ &= 0.99988 \end{aligned}$$

—

Example

Now suppose$P(A) = P(B) = P(C) = P(D) = 0.95$, calculate the reliability of the system:



$$\begin{aligned}
P(\text{System works}) &= P(\text{Subsys I works} \cap \text{Subsys II works}) \\
&= P(\text{Subsys I works})P(\text{Subsys II works}) \\
&= (1 - P(\overline{A} \cap \overline{B}))(1 - P(\overline{C} \cap \overline{D})) \\
&= (1 - 0.5^2)(1 - 0.5^2) \\
&= 0.99501
\end{aligned}$$

1.2.4 Conditional Independence

Events T_1, T_2, $\ldots$ T_n are conditionally independent if:

For all $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ and all $1 \leq k \leq n$

$$\boxed{P(T_{i_1} \cap T_{i_2} \cap \cdots \cap T_{i_k} | B) = P(T_{i_1}|B)P(T_{i_2}|B) \ldots P(T_{i_k}|B)} \tag{1.15}$$

Condition for conditional independence

Which, similarly to the condition for independence of multiple events, say where $n = 3$, the following by itself is insufficient:

$$P(T_1 \cap T_2 \cap T_3|B) = P(T_1|B)P(T_2|B)P(T_3|B)$$

The following are also required conditions:

$$P(T_1 \cap T_2|B) = P(T_1|B)P(T_2|B)$$

$$P(T_1 \cap T_3|B) = P(T_1|B)P(T_3|B)$$

$$P(T_2 \cap T_3|B) = P(T_2|B)P(T_3|B)$$

Note here that mathematically, conditional independence *does not* imply unconditional independence and vice-versa. However, in practice this can usually be assumed even though there exist counterexamples. So in this course:

If there is conditional independence on B, you may assume conditional independence on $\overline{B}$.

We will use this this assumption in developing the sequential Bayes' formula.

1.2.5 Sequential Bayes' Formula

Let S_i be the outcome of the ith test, where the outcome can either be positive or negative.

Let I_k be $S_1 \cap S_2 \cap \cdots \cap S_k$, ie. the summary of information available after test k. Also let $P(E)$ be the probability that part is positive (so the part is defective).

Then set:

$$\begin{aligned}
\pi_0 &= P(E) \\
\pi_1 &= P(E|I_1) = P(E|S_1) \\
\pi_2 &= P(E|I_2) = P(E|S_1 \cap S_2) \\
\pi_3 &= P(E|I_3) = P(E|S_1 \cap S_2 \cap S_3)
\end{aligned}$$

And so on. Assume that S_1, S_2, $\ldots$, S_n are conditionally independent on E and $\overline{E}$.

Then for $k = 1, 2, \ldots, n$:

$$\boxed{\pi_k = \frac{P(S_k|E)\pi_{k-1}}{P(S_k|E)\pi_{k-1} + P(S_k|\overline{E})(1 - \pi_{k-1})}} \tag{1.16}$$

Bayes' Updating Probability Formula

Note that this is an iterative process, whereby π_k is calculated from π_{k-1}.

In pseudo-code, the computation would look something like this:

```
S  = (1, 0, 1, 0...) // test results
n  = n               // number of tests
pi =                 // probability of a defect

pi[0] = pi
```

```
for k=1 to n:

   if S[k] == 1:
       a = p[k]
       b = q[k]

   if S[k] == 0:
       a = 1 - p[k]
       b = 1 - q[k]

   pi[k] = a*pi[k-1] / ( a*pi[k-1] + b(1-pi[k-1] )
```

// todo: insert spam example 56/61 here

1.3 Random Variables

The first point to note here is that random variables is a bad name. These are not variables, they're functions.

A random variable, denoted by an upper-case letter, is a function defined on the sample space to the reals:

$$X : \Omega \to \mathbb{R}$$

. Where the input is the resulting event from sample space, ω, and the possible value X is a real number, the lower-case letter, x.

$$X(\omega) = x$$

For instance, consider the experiment flipping a coin 10 times:

1. The sample space Ω is all possible sequences of ten heads and tails.
2. The outcome ω could be the possible result $(HTTHHTTTHT)$.
3. The random variable X could be defined as the number of heads in the outcome, and so $X(\omega) = 4$.
4. The random variable Y could be defined as the largest run of tails in the outcome, and so $Y(\omega) = 3$.

The event: "the random variable X takes the value x" is mathematically represented as $X = x$, which means $\{\omega : X(\omega) = x\}$.

This is used for probabilities, where we could say $P(x = X)$ for $x > 5$ where the event is that the outcome of the random variable X is greater than 5, and here we can compute the probability of this event.

1.3.1 Discrete Random Variables

A random variable is discrete when its range is either *finite*, such as $\{0,1,2,\ldots,100\}$, or *countable* (as in denumerable/countably infinite), such as $\{1,2,3,\ldots\}$.

1.3.1.1 Probability Mass Function (pmf)

The *pmf* of a discrete random variable X gives the probability of occurance for each possible value x of X. Denoted by a lower-case function letter.

$$\boxed{f(x) = P(X = x)} \tag{1.17}$$

pmf

Properties of the pmf:

1. $0 \leq f(x) \leq 1$
2. $\sum f(x) = 1$
3. $P(X \in A) = \sum x \in A f(x)$ — if X is a subset of real numbers, then the probability of that equals the sum of those $f(x)$

1.3.1.2 Cumulative Distribution Function (cdf)

Usually just called the distribution function, this is the probability that a random value is less or equal to a given value, denoted by an upper-case function letter.

$$\boxed{F(x) = P(X \leq x) = \sum_{k \leq x} f(k)} \tag{1.18}$$

cdf

1. $0 \leq F(x) \leq 1$
2. $F(x)$ is non-decreasing
3. $F(-\infty) = 0$, $F(\infty = 1)$
4. $P(a < X \leq b) = F(b) - F(a)$
5. $f(k) = F(k) - F(k-1)$

Example

Say the following cdf and pmf are given by the table:

x	$f(x)$	$F(x)$
0	0.15	0.15
1	0.25	0.40
2	0.30	0.70
3	0.20	0.90
4	0.10	1.00

What is the value for $P(1 < X \leq 3)$? $P(1 \leq X < 3)$?

—

$$P(1 < X \leq 3) = F(3) - F(1) = 0.90 - 0.40 = 0.50$$

With, $P(1 \leq X < 3)$, note that our cdf properties have the $\leq$ as the upper-bound. But since we know these values are discrete, we can convert the inclusion-exclusion form to an exclusion-inclusion form:

$$P(1 \leq X < 3) = P(0 < X \leq 2) = F(2) - F(0) = 0.70 - 0.15 = 0.55$$

//todo: example 20/97, and 21/97

1.3.1.3 Expected Value

This is another kind of misleading name, really this is the mean. For instance, if our options are {0,1,2,3,4} and they are all equally possible, the expected value is 3.5 but we will never actually expect this value.

The expected value of a random variable X is denoted as $E(g(X))$, where it is the weighted average of the function $g(X)$.

$$E(g(X)) = \sum_x g(x)f(x)$$

The more likely values of $g(x)$ have larger values of $f(x)$ and thus have more weight. Since $E(g(x))$ is considered a "typical value" of $g(x)$, it can be used to summarized $g(X)$.

1.3.1.4 Law of Large Numbers

The law of large numbers merely states that as the number of events increases to infinity, the average value goes to the expected value.

$$\boxed{\text{As } n \to \infty, \qquad \text{Avg}(X) = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \to E(X)} \tag{1.19}$$

Law of Large Numbers

This should be fairly obvious. The longer you flip a coin, the more likely it is the number of heads will be approximately equal to the number of tails.

1.3.1.5 Linear Operator

The expected value operator E is a linear operator, meaning that it has a distributive quality, whereby:

$$E(X + Y) = E(X) + E(Y)$$

This means that it does not depend on the probability distributions for X and Y.

Additionally, constants can be factored out of the operator, and therefore the E of a constant is just the constant:

$$\begin{aligned} E(a+bX) &= E(a)+E(bX) \\ &= aE(1)+bE(X) \\ &= a+bE(X) \end{aligned}$$

1.3.1.6 Moments

1.3.1.7 Mean, Variance, Standard Deviation

Mean:

$$\boxed{\mu = E(X) = \sum x f(x)} \tag{1.20}$$

Mean from pmf

Variance:

$$\boxed{\sigma^2 = \mathrm{Var}(X) = E\big((X-\mu)^2\big) = \sum (x-\mu)^2 f(x)} \tag{1.21}$$

Variance from pmf

Standard deviation:

$$\sigma = \sqrt{\sum (x-\mu)^2 f(x)}$$

When calculating the variance by hand, you can use:

$$\boxed{\sigma^2 = E(X^2) - E(X)^2 = \mu_2 - \mu^2} \tag{1.22}$$

Variance approximation

Example

Take the following table:

Calculate the mean, standard deviation, and variance.

—

x	$f(x)$	$xf(x)$	$x^2f(x)$
0	0.15	0.00	0.00
1	0.25	0.25	0.25
2	0.30	0.60	1.20
3	0.20	0.60	1.80
4	0.10	0.40	1.60

Mean:

$$\mu = E(X) = \sum xf(x) = 0.25 + 0.60 + 0.60 + 0.40 = \boxed{1.85}$$

$$\mu_2 = \sum x^2 f(x) = 0.25 + 1.20 + 1.80 + 1.60 = 4.85$$

Variance:

$$\sigma^2 = \mathrm{Var}(X) = \mu_2 - \mu^2 = 4.85 - 1.85^2 = \boxed{1.4275}$$

Standard deviation:

$$\sigma = \sqrt{1.4275} = \boxed{1.1948}$$

—

Example

For $x = 1, 2, \ldots, 10$:

$$f(x) = \frac{1}{2.928968} \times \frac{1}{x}$$

Calculate the mean, standard deviation, and variance.

—

$$\mu = \sum xf(x) = \sum_{x=1}^{10} x \frac{1}{2.928968x} = \sum_{x=1}^{10} \frac{1}{2.928968} = \frac{10}{2.928968} = \boxed{3.4142}$$

$$\mu_2 = \sum x^2 f(x) = \sum_{x=1}^{10} \frac{x}{2.928968} = \frac{55}{2.928968} = \boxed{18.77855}$$

$$\sigma^2 = \mu_2 - \mu^2 = 18.77855 - 3.4142^2 = \boxed{7.1218}$$

$$\sigma = \boxed{2.6687}$$

Example

Where the pmf is:

$$f(x) = \begin{cases} (1-p)^2 & x = 0 \\ 2p(1-p)^2 & x = 1 \\ p^2 & x = 2 \end{cases}$$

Calculate the expected value, variance, and standard deviation.

—

$$\mu = \sum xf(x) = 0(1-p)^2 + 1(2p(1-p)) + 2(p^2) = 2p(1-p) + 2p^2 = \boxed{2p}$$

$$\mu_2 = \sum x^2 f(x) = 0(1-p)^2 + 1(2p(1-p)) + 4(p^2) = 2p(1-p) + 4p^2 = 2p + 2p^2$$

$$\sigma^2 = \mu_2 - \mu^2 = (2p + 2p^2) - (2p)^2 = \boxed{2p - 2p^2}$$

$$\sigma = \boxed{\sqrt{2p - 2p^2}}$$

—

Example

An urn contains n chips numbered 1 through n. We draw k chips (1¡k¡n) without replacement. Let Y represent the highest number among those drawn.

(a) What is the range of Y?

(b) Find $F_Y(y)$

(c) Find $f_Y(y)$

(d) Suppose $n = 20$ and $k = 5$. Calculate the mean, variance, and standard deviation for Y.

—

(a) Suppose the first selection was the smallest possible value, 1. Then the value of Y would be 1. Then for the second selection, the next smallest value would be 2. This would continue, and so because there is no replacement, the smallest possible value for Y is the kth chip. The largest possible selection is n. So the range is $\boxed{\{k, k+1, \ldots, n\}}$.

(b) $F(y)$ is

(c) $f(y)$ is

(d)

//todo

1.3.1.8 Properties of the Mean

The mean minimizes the Mean Square Error:

$$S(t) = E[X - t)^2] \geq E[(X - \mu)^2] = \text{Var}(X), \quad \text{for all } t$$

Proof. Recall that $E(a + bX) = a + bE(X)$ for constants a, b.

$$\begin{aligned} S(t) &= E(X^2 + t^2 - 2Xt) \\ &= E(X^2) + t^2 - 2\mu t \\ S'(t) &= 2t - 2\mu = 0 \implies t = \mu \\ S''(\mu) &= 2 > 0 \quad (\mu \text{ is a minimizer}) \end{aligned}$$

□

1.3.1.9 Properties of the Variance

$$\begin{aligned} Var(a + bX) &= E[a + bX - E(a + bX)]^2 \\ &= E[a + bX - a - E(bX)]^2 \\ &= E[bX - bE(X)]^2 \\ &= E[b^2(X - E(X))^2] \\ &= b^2 E[(X - E(X))^2] \\ &= b^2 Var(X) \end{aligned}$$

Therefore:

$$\boxed{Var(a + bX) = b^2 Var(X)} \tag{1.23}$$

Property of Variance

Additionally:

$$\text{StdDev}(a + bX) = |b|\text{StdDev}(X)$$

1.3.1.10 Binomial Random Variables

Let A be an occurrence we want to monitor. We'll call the occurrence of A a success, and the non-occurrence a failure, and $p = P(A)$ where p is the probability of success

The number of instances of A that are monitored (*independent* trials) is n, where the outcome could either be a success or a failure. The number of successes is:

$$X \sim Bin(n, p)$$

The range of X is zero to the number of possible successes:

$$\text{Range} = \{0, 1, 2, \ldots, n\}$$

The density of a **binomial random variable** is (assumes independence):

$$\boxed{P(X = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n} \tag{1.24}$$

cdf of a Binomial Random Variable

Recall where:

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

This is a random variable, so it has a moment-generating function (??? get clarity of moment generating functions):

$$M(t) = (1 - p + pe^t)^n$$

$$M'(t) = \frac{\partial}{\partial t} M(t) = n(1 - p + pe^t)^{n-1} pe^t$$

$$M''(t) = n(n-1)(1-p+pe^t)^{n-1}p^2e^{2t} + n(1-p+pe^t)^{n-1}pe^t$$

The mean, $\mu = M'(0) = n(1-p+pe^0)^{n-1}pe^0 = np$, and so the mean is the number of trials multiplied by the probability of success:

$$\boxed{\mu = M'(0) = E(X) = np} \tag{1.25}$$

Mean of a Binomial Random Variable

And:

$$\mu_2 = n(n-1)p^2 + np$$

The variance is $\sigma^2 = M''(0) - M'(0)^2 = \mu_2 - \mu^2$:

$$\boxed{\sigma^2 = np(1-p)} \tag{1.26}$$

Variance of a Binomial Random Variable

Note that the variance is maximized where $p = 0.5$, because if $p = 0.9$, there are mostly successes and if $p = 0.1$, there are mostly failures.

Example

Suppose that finding oil when digging at certain locations has a probability $p = 0.10$ (geologically determined locations).

(a) How many wells should we dig to find oil with probability larger than or equal to 0.95?

(b) How many wells should we dig to obtain at least 2 successful wells with probability larger than or equal to 0.95?

—

Assume that each well is independent, then $X \sim Bin(n, 0.10)$ is the number of wells with oil where n is the number of dug wells.

(a) The probability of finding oil is $P(X > 0)$.

$$\begin{aligned} P(X > 0) &= 1 - P(X = 0) \\ &= 1 - \binom{n}{0} p^0 (1-p)^{n-0} \\ &= 1 - (1 - 0.10)^n \\ &= 0.95 \end{aligned}$$

Solving for $n \to n = 28.43$, so must dig at least $\boxed{29}$ wells.

(b)

$$\begin{aligned} 0.95 &= P(X \geq 2) \\ &= P(X > 1) \\ &= 1 - P(X = 1) - P(X = 0) \\ &= 1 - \left(\binom{n}{1} p^1 (1-p)^{n-1}\right) - \left(\binom{n}{0} p^0 (1-p)^{n-0}\right) \\ &= 1 - \left(n0.10(1 - 0.10)^{n-1}\right) - (1 - 0.10)^n \end{aligned}$$

Solving for $n \to \boxed{n = 49}$.

1.3.1.11 Poisson Random Variables

Let's say we want to count the number of occurrences of certain event A, that has a typical rate of occurrences per time period (or per area), λ. The quantity of interest is X, the number of occurrences.

$$X \sim \mathcal{P}(\lambda)$$

For a **Poisson random variable** is, the probability of k occurrences in the interval [0,t] is written as:

$$P(k; t)$$

With Poisson random variables we make the following assumptions:

1. Occurrences in disjoint time intervals are independent
2. There is **proportionality**, meaning:

$$P(1; t) = \lambda t + o(t), \quad \text{where} \lim_{t \to 0} \frac{o(t)}{t} = 0$$

Here, $o(t)$ represents the order, and this means that $o(t)$ goes to zero faster than t. (??? what's the significance of this?)

3. Events are **rare events**, meaning that events are sparse and have at most 1 occurrence of A in a small period of time:

$$1 - P(0;t) - P(1;t) = \sum_{k=2}^{\infty} P(k;t) = o(t)$$

For Poisson, X represents the number of occurrences over the range $\{0, 1, 2, \dots\}$

The pmf of a Poisson:

$$\boxed{f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \dots} \tag{1.27}$$

pmf of Poisson Random Variable

The moment generating function:

$$M(t) = e^{\lambda(e^t - 1)}$$

The mean:

$$\boxed{\mu = M'(0) = E(X) = \lambda} \tag{1.28}$$

Mean of Poisson Random Variable

And the variance:

$$\boxed{\sigma^2 = M''(0) - M'(0)^2 = Var(X) = \lambda} \tag{1.29}$$

Variance of Poisson Random Variable

Note that with a Poisson random variable, $\mu = \sigma^2 = \lambda$. When looking at real data, if your mean and variance are very close and events can be considered rare, you might be able to model your system with the Poisson distribution – if not, it could be overly or underly spaced and you cannot use Poisson.

Example

Suppose that the number of earthquakes over 5.0 on the Richter scale in a given area is a Poisson random variable and there are on average 3.6 of these earthquakes per year.

(a) What is the probability of having at least 2 earthquakes over 5.0 during the next 6 months?

(b) What is the probability of having 1 earthquake over 5.0 next month?

(c) What is the probability of waiting more than 3 months for the next earthquake over 5.0?

—

We know:

$$Y \sim \mathcal{P}(\lambda)$$

And $\lambda = 3.6/\text{year} = 0.3/\text{month} = 1.8/(6\text{-months}) = 0.9/(3\text{-months})$

(a) The probability of having at least two earth quakes in the next six months, here we use the $\lambda = 1.8/(6\text{-months})$ to keep the interval consistent.

$$\begin{aligned} P(X \geq 2) &= P(X > 1) \\ &= 1 - P(X = 1) - P(X = 0) \\ &= 1 - \frac{e^{-\lambda}\lambda^1}{1!} - \frac{e^{-\lambda}\lambda^0}{0!} \\ &= 1 - 1.8e^{-1.8} - e^{-1.8} = \boxed{0.53716} \end{aligned}$$

(b) Here we use $\lambda = 0.3/\text{month}$.

$$P(X = 1) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-0.3}0.3^1}{1!} = \boxed{0.22225}$$

(c) The probability of waiting more than three months is the same as the probability that there are no earthquakes in a rate of # earthquakes per three months, so $\lambda = 0.9/\text{month}$.

$$\begin{aligned} P(\text{Waiting ¿ 3 months}) &= P(X = 0) \\ &= \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \frac{e^{-0.9}0.9^0}{0!} = \boxed{0.40657} \end{aligned}$$

Poisson random variables are memoryless, so it doesn't matter where we start. The probability that no earthquakes happen in the next three months is the same as the probability that no earthquakes happen in the three months after that, and so on.

1.3.2 Continuous Random Variables

A random variable is continous when its range is an interval, such as (0,1).

1.3.2.1 PDF and CDF

Some properties of the density of a continuous random variable are the following:

1. The function is non-negative:
$$f(x) \geq 0$$
2. The function integrates to one:
$$\int_{-\infty}^{\infty} f(x)dx = 1$$
3. The integral of a range can be used to compete probabilities:
$$P(a < X < b) = \int_{a}^{b} f(x)dx$$

Some properties of the distribution function of a continuous random variable are the following:

1. The probability of a value greater than a given x is the cdf:
$$\boxed{F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt} \tag{1.30}$$

cdf of continuous random variable

The cdf can be used to compute the probability of a range $P(a < X < b)$. Because $\int_a^b f(x)dx = \int_{-\infty}^b f(x)dx - \int_a^\infty f(x)dx$, we can show:

$$\boxed{P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)} \tag{1.31}$$

Probability of range of continuous random variable

Here the probability is the difference between the area under the upper limit of the curve minus the area under the lower limit of the curve.

Note that it doesn't matter whether the inequalities are $<$ or $\leq$ because as equation (1.32) shows, $P(X = a) = P(X = b) = 0$, so:

$$P(a < X < b) = P(a \leq X \leq b)$$

Note: Although for discrete random variables, we had $f(x) = P(X = x)$, but this is not true of continuous random variables. For continuous random variables:

$$\boxed{P(X = x) = \int_x^x f(t)dt = 0, \qquad \text{for all x}} \tag{1.32}$$

Probability of exact value with continuous random variable

And so $f(x) \neq P(X = x)$. In fact, we often have $f(x) > 1$. However, for a very small $\delta > 0$, we can approximate:

$$P(x < X < x + \delta) = \int_x^{x+\delta} f(t)dt \approx f(x)\delta$$

Also, note that:

$$\boxed{F'(x) = \frac{\partial}{\partial x} \int_{-\infty}^x f(t)dt = f(x)} \tag{1.33}$$

pdf from cdf of continuous random variable

1.3.2.2 Mean and Variance

$$\boxed{\mu = \int_{-\infty}^{\infty} x f(x) dx} \tag{1.34}$$

Mean of continuous random variable

$$\boxed{\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2 = \mu_2 - \mu^2} \tag{1.35}$$

Variance of continuous random variable

1.3.3 Uniformly Distributed Random Variables

Uniform random variables are probably the simplest there are. Here, everything has the same density on a given interval, so each value in the interval has the same probability. This is very powerful in simulation.

The notation for a uniform random variable, where α is the lower limit and β is the upper limit, is:

$$X \sim Unif(\alpha, \beta)$$

1.3.3.1 PDF and CDF

The probability density function of a uniform random variable visually looks like a horizontal line, and is defined as:

$$\boxed{f(x) = \begin{cases} 0 & x \leq \alpha \\ (\beta - \alpha)^{-1} & \alpha < x < \beta \\ 0 & x \geq \beta \end{cases}} \tag{1.36}$$

pdf of uniform random variable

And the distribution function visually looks like a linear increase from α to β, and is 0 before and 1 after. It is defined as:

$$\boxed{F(x) = \int_{\alpha}^{x} f(t)dt = \begin{cases} 0 & x \leq \alpha \\ (x-\alpha)/(\beta-\alpha) & \alpha < x < \beta \\ 1 & x \geq \beta \end{cases}} \tag{1.37}$$

cdf of uniform random variable

But from now on we'll ignore the domain $x \leq \alpha$ and $x \geq \beta$, and so:

$$\boxed{P(a < X < b) = F(b) - F(a) = \frac{b-a}{\beta-\alpha}} \tag{1.38}$$

Probability of range of uniform random variable

When you are told for example that $X \sim \text{Unif}(0,1)$, this means:

$$f_X(x) = 1 \qquad \text{for } 0 \leq x \leq 1$$

$$F_X(x) = x \qquad \text{for } 0 \leq x \leq 1$$

1.3.3.2 Mean and Variance

The mean of a uniform random variable is simply the average of the upper and lower limit:

$$\boxed{\mu = \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} x dx = \frac{\alpha+\beta}{2}} \tag{1.39}$$

Mean of a uniform random variable

And the second mean is:

$$\mu_2 = \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} x^2 dx = \frac{\beta^2+\alpha^2+\alpha\beta}{3}$$

And so the variance is:

$$\boxed{\sigma^2 = \mu_2 - \mu^2 = \frac{(\beta-\alpha)^2}{12}} \tag{1.40}$$

Variance of uniform random variable

Example

Suppose that X is Unif(0, 10). Calculate $P(X > 3)$ and $P(X > 5|X > 2)$.

—

So $\alpha = 0$, $\beta = 10$, and $x = 3$.

$$\begin{aligned} P(X > 3) &= 1 - F(3) \\ &= 1 - \frac{x - \alpha}{\beta - \alpha} \\ &= 1 - \frac{3 - 0}{10 - 0} = \boxed{0.7} \end{aligned}$$

Now $\alpha = 0$, $\beta = 10$, and $x = 5$.

$$\begin{aligned} P(X > 5|X > 2) &= \frac{P(X > 5 \cap X > 2)}{P(X > 2)} \\ &= \frac{P(X > 5)}{P(X > 2)} && \text{In order for both to be true, } X > 5 \\ &= \frac{1 - F(5)}{1 - F(2)} \\ &= \frac{1 - \frac{5-0}{10-0}}{1 - \frac{2-0}{10-0}} = \boxed{0.625} \end{aligned}$$

Example

Suppose that X is Unif(0,1). Derive the distribution function and density function for $Y = -\ln(X)$.

—

Since X is Unif(0,1), then:

$$f_X(x) = 1, \; 0 \leq x \leq 1$$

$$F_X(x) = x,\ 0 \le x \le 1$$

Notice that the range of Y is $(0, \infty)$, so for $y < 0$:

$$F_Y(y) = 0$$

—

For $y > 0$:

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(-\ln(X) \le y) \\ &= P(X \ge e^{-y}) \\ &= 1 - F_X(e^{-y}) \\ &= \boxed{1 - e^{-y}} \end{aligned} \tag{1.41}$$

Similarly, for $y > 0$:

$$f_Y(y) = F_Y'(y) = \frac{\partial}{\partial y}(1 - e^{-y}) = \boxed{e^{-y}}$$

1.3.4 Exponential Random Variables

Exponential random variables are used to model the waiting time until the occurrence of a certain event. Exponential density functions have the input rate of occurrence, $\lambda = \#$ of occurrences/time intervals, and the notation for an exponential random variable is:

$$X \sim Exp(\lambda)$$

1.3.4.1 PDF and CDF

The density function for an exponential random variable is:

$$\boxed{f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}} \tag{1.42}$$

pdf for exponential random variable

And similarly for $x \leq 0$, $F(x) = 0$. For $x > 0$, the distribution function is:

$$\boxed{P(X \leq x) = F(x) = \int_0^x e^{-\lambda t} dt = 1 - e^{-\lambda x}} \tag{1.43}$$

cdf for exponential random variable

1.3.4.2 Mean and Variance

The mean of an exponential random variable is:

$$\boxed{\mu = \frac{1}{\lambda}} \tag{1.44}$$

Mean of exponential random variable

And the variance:

$$\boxed{\sigma^2 = \frac{1}{\lambda^2}} \tag{1.45}$$

Variance of exponential random variable

Note that:

$$\boxed{Exp(\lambda) = \frac{-\ln\left(Unif(0,1)\right)}{\lambda}} \tag{1.46}$$

Exponential from uniform random variable

Note that exponential random variables are memoryless, so they have a constant probability. Where $X \sim Exp(\lambda)$:

$$P(X > h) = P(X > h + x \mid X > x) = e^{-\lambda h}$$

So at age x, the probability of surviving h additional units is the same for all ages x.

1.3.4.3 Failure Rate

$$\boxed{\lambda(x) = \frac{f(x)}{1 - F(x)} = -\frac{\partial}{\partial x}\ln[1 - F(x)]} \tag{1.47}$$

Failure rate

If the failure rate increases, it means the system wears out over time. Exponential random variables are the only type of random variable in the continuous domain that have a constant failure rate.

$$\lambda(x)\delta \approx P(X \leq x + \delta \mid X > x), \qquad \text{for a small } \delta$$

Given the failure rate, from (1.47), we can calculate the distribution function:

$$\boxed{F(x) = 1 - e^{-\int_0^x \lambda(t)dt}} \tag{1.48}$$

cdf from failure rate

For a constant failure rate, where $\lambda(x) = k$:

$$F(x) = 1 - e^{-kx}$$

An increasing failure rate is often used in engineering. For instance, the Weibull distribution is where $\lambda(x) = x$ and is often used to demonstrate where there is a linear increase in failure rate:

$$\boxed{F(x) = 1 - e^{-\frac{1}{2}x^2}} \tag{1.49}$$

Weibull distribution

For a decreasing failure rate, for example where $\lambda(x) = \frac{1}{1+x}$:

$$F(x) = 1 - e^{-\int_0^x \frac{1}{1+t}dt} = 1 - e^{-[\ln(x+1) - \ln(1)]} = 1 - \frac{1}{x+1} = \frac{x}{x+1}$$

1.3.5 Normal Distribution

1.3.5.1 Standard Normal Distribution

The standard normal random variable is denoted by Z. The standard normal density is where $\mu = 0$ and $\sigma^2 = 1$:

$$\boxed{\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}} \tag{1.50}$$

Standard normal density function

Standard normal distribution function:

$$\boxed{P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \varphi(t)dt} \tag{1.51}$$

Standard normal distribution function

$\Phi(z)$ cannot be calculated in closed form, meaning it can't be calculated in a finite number of operations, so you'll usually either look it up in a table or use a computer, such `pnorm(z)` in R. The normal density function is symmetrical, therefore:

$$\boxed{\Phi(z) = 1 - \Phi(-z)} \tag{1.52}$$

Symmetry formula

$$\boxed{\mu = E(Z) = \int_{-\infty}^{\infty} z\varphi(z)dz = 0} \tag{1.53}$$

Mean of standard normal distribution

$$\boxed{\sigma^2 = Var(Z) = \int_{-\infty}^{\infty} z^2\varphi(z)dz = 1} \tag{1.54}$$

Variance of standard normal distribution

1.3.5.2 General Normal Random Variables

The notation for a normal random variable is:

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

You can create a non-standard normal distribution with measurements X_1, X_2, ..., X_n from a standard normal distribution Z_1, Z_2, ..., Z_n, where:

$$X_i = \mu + \sigma Z_i \qquad i = 1, 2, \ldots, n$$

And because $E(Z) = 0$ and $Var(Z) = 1$ for the standard normal variable Z:

$$\boxed{E(X) = E(\mu + \sigma Z) = \mu + \cancelto{0}{\sigma E(Z)} = \mu}$$

$$\boxed{Var(X) = Var(\mu + \sigma Z) = \sigma^2 \cancelto{1}{Var(Z)} = \sigma^2}$$

The distribution function for X is:

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \end{aligned}$$

$$\boxed{F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)} \tag{1.55}$$

cdf of normal random variable

The density function for X is:

$$\boxed{f(x) = F'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}} \tag{1.56}$$

pdf of normal random variable

Example

Let $Z \sim N(0,1)$ and calculate:

(a) $P(0.10 \leq Z \leq 0.35)$

(b) $P(Z > 1.25)$

(c) $P(Z > -1.20)$

(d) Find c such that $P(Z > c) = 0.05$

(e) Find c such that $P(|Z| < c) = 0.95$

—

(a)

$$P(0.10 \leq Z \leq 0.35) = \Phi(0.35) - \Phi(0.10) = \boxed{0.0970}$$

(b)

$$\begin{aligned} P(Z > 1.25) &= 1 - P(Z \leq 1.25) \\ &= 1 - \Phi(1.25) = \boxed{0.1056} \end{aligned}$$

(c)

$$\begin{aligned} P(Z > -1.20) &= 1 - P(Z \leq -1.20) \\ &= 1 - (1 - \Phi(1.20)) = \boxed{0.8849} \end{aligned}$$

(d) Find c such that $P(Z > c) = 0.05$

$$\begin{aligned} P(Z > c) &= 0.5 \\ 1 - P(Z \leq c) &= 0.5 \\ P(Z \leq c) &= 0.95 \\ \Phi(c) &= 0.95 \\ c &= \Phi^{-1}(0.95) = \boxed{1.644854} \end{aligned}$$

(e) Find c such that $P(|Z| < c) = 0.95$

$$\begin{aligned} P(-c < Z < c) &= 0.95 \\ \Phi(c) - \Phi(-c) &= 0.95 \\ \Phi(c) - (1 - \Phi(c)) &= 0.95 \\ 2\Phi(c) &= 1.95 \\ c = \Phi^{-1}\left(\frac{1.95}{2}\right) &= \boxed{1.959964} \end{aligned}$$

1.3.5.3 Summary

Table 1.1: Summary of Probability Distribution Models

What we know	**What to determine**	**What model to use**
Occurrences with a probability of success	Number of successes	Binomial
Rate of occurrence	Number of occurrences	Poisson
		Uniform
Rate of occurrence	Waiting time until occurrence	Exponential

**** WHOLE BUNCH OF MISSING STUFF NOT INCLUDED ON MIDTERM ****

Chapter 2

Stochastic Signals and Systems Applied to EE

Construction of the event-space definition is the most challenging task, both in industry and in academia.

Independence is not pair-wise independence.

Set's say with have A given B. Let's say you're trying to estimate impedance, your random variable A

For electrical systems it's a Gaussian noise system. When you get a current it's billions and trillions of random electron movements, and we just average it all out.

Appendix

A Tutorial 1 (Lena)

Example 1

Where A and B are events and:

$$P(A) = \frac{3}{4};\ P(B) = \frac{1}{3}$$

Show:

$$\frac{1}{12} \leq P(A \cap B) \leq \frac{1}{3}$$

—

First thing to note is that the intersection cannot be larger than the smallest P, so this is the upper bound:

$$P(A \cap B) \leq \frac{1}{3}$$

From the inclusion-exclusion, $P(A \cap B) = P(A) + P(B) - P(A \cup B)$. Since $0 \leq P(A \cup B) \leq 1$, then the minimum bound is where $P(A \cup B) = 1$ and:

$$P(A \cap B) \geq P(A) + P(B) - 1 = \frac{1}{3} + \frac{3}{4} - 1 = \frac{1}{12}$$

Example 2

Prove:

$$P(\cup_{n=1}^{k} A_n) \leq \sum_{n=1}^{n} P(A_n)$$

—

Proof. Consider the base case, $n = 2$. By the inclusion-exclusion principle:

$$P(\cup_{i=1}^{2} A_i) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Since $P(A_1 \cap A_2) \geq 0$, then

$$P(\cup_{i=1}^{2} A_i) \leq P(A_1) + P(A_2) = \sum_{i=1}^{2} P(A_i)$$

Induction hypothesis, for $k \in \mathbb{N}$:

$$P(\cup_{i=1}^{k} A_i) \leq \sum_{i=1}^{k} P(A_i)$$

For $k + 1$, by the inclusion-exclusion principle:

$$\begin{gathered} P(\cup_{i=1}^{k+1} A_i) = P(\cup_{i=1}^{k} A_i \cup A_{k+i}) \leq P(\cup_{i=1}^{k} A_i) + P(A_{k+1}) \\ \leq \sum_{i=1}^{k} P(A_i) + P(A_{k+1}) = \sum_{i=1}^{k+1} P(A_i) \end{gathered}$$

□

Problem 1.1

(a) $P(\overline{B}) = 1 - P(B) = \frac{2}{3}$

(b) $P(A \cup \overline{B}) = P(A) + P(\overline{B}) - P(A \cap \overline{B}) \to P(A) = P(A \cap B) + P(A \cap \overline{B})$

(c) $P(\overline{A} \cap B) \to P(B) = P(A \cap B) + P(\overline{B} \cap \overline{A})$

(d) $P(\overline{A} \cup B) = P(\overline{A}) + P(B) - P(\overline{A} \cap B)$

Problem 1.3

A, B, and C are mutually exclusive, so they are not independent. Additionally, they have no intersections.

(a) $P(A \cup B \cup C) = P(A) + P(B) + P(C)$

(b) $P(\overline{A} \cup \overline{B}) = 1$

(c) $P(\overline{A} \cap \overline{B}) = 1 - P(A) - P(B)$

1.4

(a) $P(A_{good}) = \frac{3}{5}$

(b)

$$P(B_{good}|A_{good}) = \frac{P(B_{good} \cap A_{good})}{P(A_{good})} = \frac{\frac{\binom{3}{2}}{\binom{5}{2}}}{\frac{3}{5}} = \frac{\frac{3}{10}}{\frac{3}{5}} = \frac{1}{2}$$

$$P(B_{good}|A_{bad}) = \frac{P(B_{good} \cap A_{bad})}{P(A_{bad})} = \frac{1}{2} \times \frac{\frac{\binom{3}{1}\binom{2}{1}}{\binom{5}{2}}}{1 - \frac{3}{5}} = \frac{1}{2} \times \frac{\frac{3}{5}}{1 - \frac{3}{5}} = \frac{3}{4}$$

Times by 1/2 because we only want the case, B_{good} and A_{bad}.

(c)

(d)

$$P(A_{good} \cap B_{good}) = \frac{\binom{3}{2}}{\binom{5}{2}}$$

(e)

$$P(B_{good}) = P(B_{good} \cap A_{good}) + P(B_{good} \cap A_{bad}) = \frac{\binom{3}{2}}{\binom{5}{2}} + \frac{\frac{1}{2} \times \binom{3}{1}\binom{2}{1}}{\binom{5}{2}}$$

Problem 1.5

(a)
$$P(\{balls\ are\ red\}) = \frac{\binom{7}{2}}{\binom{15}{2}}$$

(b)
$$P(\{both\ same\ colour\}) = \frac{1}{\binom{15}{2}}$$

(c)

$$P(\{2nd\ is\ green\}) = P(G_2|G_1)P(G_1) + P(G_2|B_1)P(B_1) + P(G_2|R_1)P(R_1)$$

$$\frac{4}{14} + \frac{5}{15} + \frac{5}{14} \times \frac{10}{15}$$

B Tutorial 2 (Yang)

1.9

10 keys → 1 correct. No replacement.

Define $A_i = \{\ i\text{th attempt works}\ \}$

(a) $P(A_1) = \frac{1}{10}$

(b) $P(A_2) = P(A_1^c \cap A_2) = P(A_2|A_1^c)P(A_1^c) = \frac{1}{9}\frac{9}{10} = \frac{1}{10}$

Alternatively, $P(A_2) = P(A_1 \cap A_2) + P(A_1^c \cap A_2)$ where $P(A_1 \cap A_2)$ because only 1 can be correct.

(c) $P(A_i) = P(A_1^c \cap \cdots \cap A_{i-1}^c \cap A_i) = P(A_i|A_1^c \cap \cdots \cap A_{i-1}^c) \times P(A_{i-1}^c|A_{i-2}^c \cap \cdots \cap A_1^c) \ldots P(A_1^c) = 1/10$

# of attempts	1	2	3	… 10
p_i	1/10	1/10	1/10	… 1/10

(d)

$$E(x) = \sum_{i=1}^{n} x_i p_i = 1/10 + 2/10 + 3/10 + \cdots + 10/10 = 55/10 = 5.5$$

1.10

10 keys $\rightarrow$ 1 correct. Yes replacement.

Define $A_i = \{ \text{ } i\text{th attempt works } \}$

(a) $P(A_1) = \frac{1}{10}$

(b) $P(A_2) = P(A_1^c \cap A_2) = P(A_2|A_1^c)P(A_1^c) = \frac{1}{10}\frac{9}{10} = \frac{9}{100}$

(c) Probability that all attempts up to i didn't work and that i did.

$P(A_i) = P(A_1^c \cap \cdots \cap A_{i-1}^c \cap A_i) = (\frac{9}{10})^{i-1}\frac{1}{10}$

(d) $x = \#$ of first successful attempts, $p_i = P(X = x)$

x_i	1	2	3	...	∞
p_i	1/10	$(9/10)^1$ 1/10	$(9/10)^2$ 1/10	...	

$$E(x) = \sum_{i=1}^{n} x_i p_i = \sum_{i=1}^{\infty} i \left(\frac{9}{10}\right)^{i-1} \frac{1}{10}$$

If doing infinite sum, probably need to be a geometric series. Might also be telescoping sum, but most likely geometric series.

Geometric series: $\sum_{i=0}^{\infty} p^i = \frac{1}{1-p}$ where $|p| < 1$.

Differentiate w.r.t p:

$$\sum_{i=0}^{\infty} ip^{i-1} = \frac{1}{(1-p)^2}$$

$$\sum_{i=1}^{\infty} ip^{i-1} = \frac{1}{(1-p)^2}$$

Multiply both sides by $(1-p)$:

$$\sum_{i=1}^{\infty} ip^{i-1}(1-p) = \frac{1}{1-p}$$

Since $p = 9/10$:

$$\frac{1}{1-p} = 10$$

So he will likely grab it on the 10th try.

1.11

First some stuff to recall:

Sensitivity is that if the patient is sick, the test will be positive. Specificity is that if the patient is not sick, that the test will be negative.

Equation (0):

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Equation (1):

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

Equation (2):

$$P(A) = P(B|A)P(A) + P(B^c|A)P(A)$$

Equation (3):

$$P(B|A) + P(B^c|A) = 1$$

Equation (4):

Conditional independence, events A_1 and A_2 are independent on B:

$$P(A_1 \cap A_2 | B) = P(A_1|B)P(A_2|B)$$

General independence:

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

—

Given $E = \{$ patient is sick $\}$. $T_i = \{$ test result i positive $\}$, $i = 1, 2, 3, 4, 5$.

T_i are conditionally independent given E and E^c.

Known: $P(E)$, $P(T_i|E)$, $P(T_i^c|E^c)$

$$P(T_i^c|E) = 1 - P(T_i|E) \rightarrow$$

(a) $P(E|I_n)$, $I_n = T_1 \cap T2 \cap \cdots \cap T_5$. From $\frac{(1)}{(2)}$.

$$P(E|I_n) = \frac{P(E \cap I_n)}{P(I_n)} \frac{P(I_n|E)P(E)}{P(I_n \cap E) + P(I_n \cap E^c}$$

Look at notes online...

$$= \frac{P(T_1|E) \times \cdots \times P(T_n|E) \times P(E)}{P(T_1|E) \ldots}$$

(b) Sequential updates. $I_{k+1} = T_1 \cap T2 \cap ...T_k \cap T_{k+1}$.

$$P(E|I_{k+1}) = \frac{P(E \cap I_{k+1})}{P(I_{k+1})}$$

Numerator, from (1):

$$P(E \cap I_{k+1}) = P(I_{k+1} \cap E) = P(I_{k+1}|E)P(E)$$

$$= P(T_1 \cap T2 \cap ...T_k \cap T_{k+1}|E)P(E)$$

$$= P(I_k \cap T_{k+1}|E)P(E)$$

Because independent, from formula (4):

$$= P(I_k|E)P(T_{k+1}|E)P(E)$$

Unite $P(I_k|E)P(E) = P(I_k \cap E)$ from (1):

$$= P(I_k \cap E)P(T_{k+1}|E)$$

$$= P(E|I_k)P(I_k)P(T_{k+1}|E)$$

Call this result $(\star)$.

Denominator:

$$P(I_{k+1}) = P(I_{k+1} \cap E) + P(I_{k+1} \cap E^c$$

Conditioning on E^c:

$$P(I_{k+1} \cap E^c) = P(I_{k+1}|E^c)P(E^c)$$

$$= P(I_k \cap T_{k+1}|E^c)P(E^c)$$

$$= P(I_k|E^c)P(T_{k+1}|E^c)P(E^c)$$

$$= P(I_k \cap E^c)P(T_{k+1}|E^c)$$

$$= P(E^c|I_k)P(T_{k+1}|E^c)P(I_k)$$

Call this result $(\star\star)$.

Putting this all together:

$$P(E|I_{k+1}) = \frac{(\star)}{(\star) + (\star\star)} = \frac{P(E|I_k)P(I_k)P(T_{k+1}|E)}{P(E|I_k)P(I_k)P(T_{k+1}|E) + P(E^c|I_k)P(T_{k+1}|E^c)P(I_k)}$$

Canceling out $P(I_k)$, we get our update formula:

$$P(E|I_{k+1}) = \frac{(\star)}{(\star star)} = \frac{P(E|I_k)P(T_{k+1}|E)}{P(E|I_k)P(T_{k+1}|E) + P(E^c|I_k)P(T_{k+1}|E^c)}$$

(c) $P(E) = 0.005$, $P(T_i|E) = 0.99$, $P(T_i^c|E^c) = 0.99$.

Since this is our first test, we use just $P(E)$ as our history:

Test 1 negative:

$$P(E|T_1^c) = \frac{P(E)P(T_1^c|E)}{P(E)P(T_1^c|E) + P(E^c)P(T_1^c|E^c)} = \frac{0.005 \times 0.01}{0.005 \times 0.01 + (1 - 0.005) \times 0.99} = 5.1(10)^{-5}$$

Test 2 positive, so no complement over T_2. And $I_k = T_1^c$:

$$P(E|T_1^c \cap T_2) = \frac{P(E|T_1^c)P(T_2|E)}{P(E|T_1^c)P(T_2|E) + P(E^c|T_1^c)P(T_2|E^c)}$$

$$= \frac{5.1(10)^{-5} \times 0.99}{5.1(10)^{-5} \times 0.99 + (1 - 5.1(10)^{-5}) \times 0.01} = 0.005$$

Best guess is still just the incidence of disease since the two tests contradict each other and neither is weighted more than the other.

Test 3 positive:

$$P(E|T_1^c \cap T_2 \cap T_3) = \frac{P(E|T_1^c \cap T_2)P(T_3|E)}{P(E|T_1^c \cap T_2)P(T_3|E) + P(E^c|T_1^c \cap T_2)P(T_3|E^c)}$$

$$= \frac{0.005 \times 0.99}{0.005 \times 0.99 + (1 - 0.005) \times 0.01} = 0.332$$

Test 4 positive:

$$P(E|T_1^c \cap T_2 \cap T_3 \cap T_4)$$

$$= \frac{0.332 \times 0.99}{0.332 \times 0.99 + (1 - 0.332) \times 0.01} = 0.9301$$

Test 5 positive:

$$P(E|T_1^c \cap T_2 \cap T_3 \cap T_4 \cap T_5) = 0.9997$$

Aside

Difference between events and probabilities.

Events: $P(A \cup B)$ works, $P(A + B)$ doesn't.

Probabilities: $P(A) + P(B)$ works, $P(A) \cup P(B)$ doesn't. Probabilities are numbers, events aren't.

C Tutorial 3 (Lena)

C.1 Random Variables Revisited

Usually notated with an upper case letter, say X. The thing that differentiates it from an unknown x is that it can change on the fly.

It is based on the probability density function, PDF, $f_X(x)$, where the subscript is the random variable and the input is the x-axis.

Densities can take values above 1, it is not a probability, it is a measure of relative frequency. We only use this when we're plotting or integrating. Usually it's a pain otherwise.

The cumulative distribution function, CDF, $F_X(x)$ is the probability that our input variable $x \leq X$. Cumulative because it sums PDF.

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{\infty} f_x(t)dt$$

The **quantile**, q, is such a value of x that is the inverse of the CDF. Where p is the area under the $f_X(x)$ curve below $x = q$.

$$P(X \leq q) = p$$

1st quantile median.

$$P(X \leq q) = 0.25$$

$$P(X \leq q) = 0.5 = F_x(q)$$

Properties of CDFs and PDFs:

1. PDF
   (a) $f_X(x) \geq 0$
   (b) if $a \leq x \leq B$, then $f_X(x) = 0$ for $x \in (-\infty, a] \cup [B, \infty)$

(c) $\int_{-\infty}^{\infty} f_X(x) = 1$

2. CDF

   (a) $0 \leq F_X(x) \leq 1$

   (b) if $a \leq x \leq B$,then
   $F_X(B) = 1 = P(X \leq B)$ and $F_X(a) = 0 = P(X \leq a)$

   (c) $F_X(x)$ is non-decreasing in x

C.2 Uniform Distribution

When the PDF is plotted, looks like a rectangle.

$$f_X(x) = \frac{1}{b-a}, \ a \leq x \leq b$$
$$F_X(x) = \frac{x-a}{b-a}, \ a \leq x \leq b$$

Since the probability is the same everywhere, if you integrate you'll see that the expected value is just in the middle.

$$E(X) = \frac{a+b}{2}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{(b-a)^2}{12}$$

C.3 Exponential Distribution

When the PDF is plotted, it looks like a typical delay.

Usually used to model time between events. There is only one parameter λ, which is the rate per item of time or per space. X is then measured in units of whatever λ is.

$X \sim Exp(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0$$

$$F_X(x) = 1 - e^{-\lambda x}, x \geq 0$$

$$E(x) = \frac{1}{\lambda}$$

Example

$\lambda = 3/day$. X = time until next call. So $X \sim Exp(3)$.

What is the probability that the next call is within 5-6 hours?

$$P(5\ hours \leq X \leq 6\ hours)$$

$$P(5/24\ days \leq X \leq 6/24\ days)$$

$$= \int_{5/24}^{6/24} 3e^{-3x}dx = F(6/24) - F(5/24) = (1 - e^{-3(6/24)}) - (1 - e^{-3(5/24)})$$

What is the median time until the call? This means when 50% of the time is below and 50% of the time is above.

$$F_X(m) = 0.5 = 1 - e^{-3m} \rightarrow \text{solve for } m$$

Expected time for a call is $\frac{1}{\lambda}$. Variance is $\left(1/\frac{3}{24}\right)^2$

C.4 Normal Distribution

$f_X(x)$ looks like a bell-curve, symmetric around μ. Parameters are μ, the mean, and σ^2, the variance.

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(t-\mu)^2}{2\sigma^2}}$$

$$F_X(x) = \cdots = \Phi(x)$$

We can't evaluate the CDF in closed-form integration.

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

Specific case we'll often see is **standard normal**, Z. At which $Z \sim N(0,1)$, $\mu = 0$ and $\sigma^2 = 1$.

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

$$Z = \frac{X - \mu}{\sigma}$$

$$\Phi(-c) = 1 - \Phi(c)$$

$$P(|Z| \leq c) = P(-c \leq Z \leq c) = \Phi(c) - \Phi(-c) = 2\Phi(c) - 1$$

Since we can't integrate in closed-form, we'll use R functions.

CDF can be found using $\Phi(c) =$ `pnorm(c)`.

$$\Phi(2) = P(Z \leq 2) = 0.9772$$

This is the area under the curve until the point $x = 2$.

Note: Symmetry means the median is the same as the mean.

For quantities, $\Phi(c) = 0.75$, where $c =$ `qnorm(0.75)` $= 0.674$. This means that the area under the curve at the point $x = 0.674$ is 75% of the area under the curve.

Problem B.5

$Z \sim N(0,1)$, $\Phi(1) = 0.3413$.

1. (a) $P(-1 < Z < 1)$

$$P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = \Phi(1) - \Phi(-1)$$

$$= \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 = 0.6826$$

(b) $P(Z > 1)$

$$P(Z > 1) = 1 - P(Z \leq 1) = 1 - \Phi(1) = 0.1587$$

(c) $P(Z < -1)$

$$P(Z < -1) = \text{same as above by symmetry}$$

2. (a) Note there are typos here, the $P(Z < c)$ should say c_1 and c_2 respectively. c_1 is larger.

$$\Phi(c_2) = 0.8, \ c_2 = \texttt{qnorm(0.8)} = 0.84$$

$$P(|Z| < c_1) = 2\Phi(c_1) - 1 = 0.8$$

$$\Phi(c_1) = 1.8/2 = 0.9$$

$$c_1 = \texttt{qnorm(0.9)} = 1.28$$

—

Non-standard normal.

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

$X = \mu + \sigma Z$, where $Z \sim N(0, 1)$

$$P(X \leq c) = P(\frac{x - \mu}{\sigma} \leq \frac{c - \mu}{\sigma}) = P(Z \leq \frac{c - \mu}{\sigma}) = \Phi(\frac{c - \mu}{\sigma})$$

Properties $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$

- $X_1 \pm X_2, \sim N(\mu \pm \mu_2, \sigma_1^2 + \sigma_2^2)$
- $aX_1 \pm BX_2, \sim N(a\mu \pm B\mu_2, a^2\sigma_1^2 + B^2\sigma_2^2)$
  Important distinction

$$Var(aX_1) = a^2Var(X_1) = a^2\sigma^2$$

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) = \sigma_1^2 + \sigma_2^2$$

Example

Distribution of grades is normal mean 75 and standard deviation is 10. Define X as the test score of the student.

1. What is the probability that a random student passes?

$$X \sim N(75, 10^2)$$

$$P(\frac{X-75}{10} \geq \frac{50-75}{10}$$

$$P(Z \geq -2.5)$$

$$1 - P(Z < -2.5)$$

$$1 - \Phi(-2.5) = \texttt{1 - pnorm(-2.5)} = 0.993709$$

2. What is the P that the student scored higher than average but lower than 85?

$$P(75 \leq X \leq 85) = P(\frac{75-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{85-\mu}{\sigma})$$

$$= P(\frac{75-75}{10} \leq Z \leq \frac{85-75}{10}$$

$$= P(0 \leq Z \leq 1) = \Phi(1) - \Phi(0) = \texttt{pnorm(1) - 0.5} = 0.3413$$

Problem 3, Lecture 4

A machine fills "10-pound bags" of concrete mix. The actual weight is normally distributed with standard deviation of $\sigma = 0.1$ lbs. The mean is set by operator. Let X by the weight of the bag, where

$$X \sim N(\mu, 0.1^2)$$

1. What μ do we need if at most 10% of bags should be underfilled (< 10lbs)?

$$P(X < 10) \leq 0.10$$

Standardize, subtract the mean.

$$P(\frac{X-\mu}{\sigma} < \frac{10-\mu}{\sigma}) = P(Z < \frac{10-\mu}{0.1}) \leq 0.1$$

Solve $\Phi(\frac{10-\mu}{0.1}) = 0.1$

$$\texttt{qnorm(0.1)} = -1.2316 = \frac{10-\mu}{0.1} \rightarrow \mu = 10.128$$

If you increase μ, you decrease the probability that the $X < 10\%$ because you reduce the area under the curve below 0.1.

So for $P(X < 10) \leq 0.1$, we need $\mu \geq 10.128$.

2. Now define $\sigma = 0.1\mu$. Hence we have $X \sim N(\mu, 0.1^2\sigma^2)$. Same question as above.

   Standardize again:

$$P(X < 10) = P\left(\frac{X-\mu}{\sigma} < \frac{10-\mu}{\sigma}\right) = P\left(Z < \frac{10-\mu}{0.1\mu}\right) = \Phi\left(\frac{10-\mu}{0.1\mu}\right)$$

$$\frac{10-\mu}{0.1\mu} = -1.216(???) \rightarrow \mu \geq 11.47$$

C.5 Failure Rates

Easy problems, just know the definitions of CDF and PDF.

Example

Lifetime of relationship has failure rate in years:

$$\lambda(t) = \frac{1}{4}e^{-t}$$

1. Are relationships stronger or weaker over time?

Since the failure rate decreases over time, then the chance that the relationship breaks is shorter the longer the relationship is, so relationships get stronger over time.

2. Mike has just started to date Lindsey. What is the probability that their relationship survives 6-months. X = duration of relationship.

$$F_X(x) = 1 - \exp \int_0^x \lambda(t)dt$$

$$P(X > 0.5) = 1 - P(X \leq 0.5) = 1 - F_X(0.5) = 1 - (1 - \exp(-\int_0^{0.5} \frac{1}{4}e^{-t}dt))$$

$$= \exp(-\int_0^{0.5} \frac{1}{4}e^{-t}dt) = \exp(\frac{1}{4}e^{-0.5} - \frac{1}{4}) = 0.90631$$

We solved this on the fly. Don't take it as gospel.

3. Find the median lifetime $F_X(m) = 0.5$.

$$1 - \exp(-\int_0^m \lambda(t)dt) = 0.5$$

D Midterm Review Session

Let's take a look at the moment generating function for a binomial distribution,

$$Y \sim Bin(n, p)$$

.

You should know, the moment generating function is:

$$M(t) = (1 - p + pe^t)$$

To get the expected value, take:

$$E(X) = M'(0) = \frac{\partial}{\partial t}M(t) = n(1 - p + pe^t)^{n-1}pe^t = np$$

Using two random variables, $X \sim B(n_1, p)$ and $Y \sim B(n_2, p)$. X is number of successes in n_1 trials and Y is number of successes in n_2 trials.

Number of successes in $n_1 + n_2$ trials is $(n_1 + n_2)p$.

Using moment generating function,

$$S = X + Y$$

$$M_s(t) = E(e^{t(X+Y)}) = E(e^{tx}e^{ty}) = (1 - p + pe^t)^{n_1}(1 - p + pe^t)^{n_2}$$

$$M_s'(0) = (n_1 + n_2)p$$

Some probability formulas:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = P(A \cap B)/P(B)$$

$$P(A \cap B) = P(A)P(B|A)$$

$$P(A) = P(A \cap B) + P(A \cap \overline{B})$$

Independence:

$$P(A \cap B) = P(A)P(B)$$

Reliability of system. In series, and. In parallel, minus and with complements.

Given non-standard normal, turn into standard using $x - \mu/\sigma$.

$$0.90 = P((X - \mu)/\sigma < 0.5/\sigma) = P(|Z| < 0.5/\sigma)$$

Failure rate:

$$\lambda(x) = 3x^2$$

$$F(x) = 1 - e^{-\int_0^x (3t^2)dt} = 1 - e^{-x^3}$$

$$P(X > 1)$$