ELEC/STAT 321

Stochastic Signals and Systems

Charles Clayton

April 18, 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.

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

Simple Form of Bayes' Formula

Generalized, the formula looks like so:

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

General Form of Bayes' Formula

Where D_1, D_2 ... 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 probabil-

ity 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:

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

Condition for independence

If A and B are independent, then:

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

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 $1 \leq k \leq n$

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

$$\rightarrow \boxed{a} \rightarrow \boxed{b} \rightarrow \boxed{c} \rightarrow$$
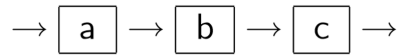
$$P(\text{System works}) = P(A \cap B \cap C) = P(A)P(B)P(C) = 0.95^3 = 0.857$$

Now calculate the reliability of the system:
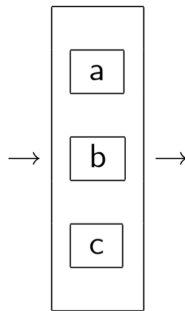
$$
\begin{aligned}
P(\text{System works}) &= 1 - P(\text{System fails}) \\
&= 1 - P(\overline{A} \cap \overline{B} \cap \overline{C}) \\
&= 1 - P(\overline{A})P(\overline{B})P(\overline{C}) \\
&= 1 - (1 - 0.95)^3 \\
&= 0.99988
\end{aligned}
$$

Example

Now suppose $P(A) = P(B) = P(C) = P(D) = 0.95$, calculate the reliability of the system:
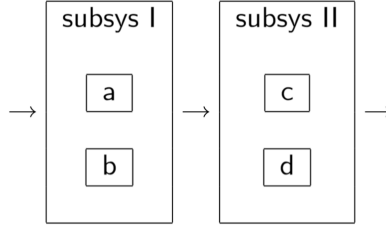
$$P(\text{System works}) = P(\text{Subsys I works} \cap \text{Subsys II works})$$
$$= P(\text{Subsys I works})P(\text{Subsys II works})$$
$$= (1 - P(\overline{A} \cap \overline{B}))(1 - P(\overline{C} \cap \overline{D}))$$
$$= (1 - 0.5^2)(1 - 0.5^2)$$
$$= 0.99501$$

1.2.4   Conditional Independence

Events T_1, T_2, ... T_n are conditionally independent if:

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

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

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, ..., S_n are conditionally independent on E and $\overline{E}$.

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

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

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.

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

15

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.

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

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:

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

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 $(1 \le k \le n)$ 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.

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

$\square$

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:

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

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

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

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

22

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

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)

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

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:

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

24

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 $\lambda = 0.9/\text{month}$.

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

26

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:

$$\boxed{F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt} \qquad (1.30)$$

cdf of continuous random variable

27

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:

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

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:

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

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:

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

pdf from cdf of continuous random variable

1.3.2.2   Mean and Variance

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

Mean of continuous random variable

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

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:

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

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:

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

cdf of uniform random variable

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

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

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:

$$\mu = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x dx = \frac{\alpha + \beta}{2} \qquad (1.39)$$

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:

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

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 $x = 3$.

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

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

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

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:

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

31

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

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

$$F_Y(y) = 0$$

For $y > 0$:

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

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

32

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

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

(1.43)

cdf for exponential random variable

1.3.4.2 Mean and Variance

The mean of an exponential random variable is:

$$\mu = \frac{1}{\lambda}$$

(1.44)

Mean of exponential random variable

And the variance:

$$\sigma^2 = \frac{1}{\lambda^2}$$

(1.45)

Variance of exponential random variable

Note that:

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

(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

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

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:

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

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:

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

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 $\sigma^2 = 1$:

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

Standard normal density function

Standard normal distribution function:

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

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:

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

Symmetry formula

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

Mean of standard normal distribution

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

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:

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

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

The distribution function for X is:

$$F(x) = P(X \le x)$$

$$= P\left(\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}\right)$$

$$= P\left(Z \le \frac{x - \mu}{\sigma}\right)$$

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

cdf of normal random variable

The density function for X is:

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

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)

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

(c)

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

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

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

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

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

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

1.4 Random Vectors

1.4.1 Continuous Random Vectors

X_1 and X_2 are random variables, that may or may not be independent.

$$\text{Random vector, } X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

Random vectors also have pdf, called *joint pdf* or *joint density*.

$$f_{X1,X2}(x1, x2)$$

Properties of joint pdf:

- Non-negative – $f_{X1,X2}(x1, x2) \geq 0$ for all $x_1 \in R$ and $x_2 \in R$

-
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X1,X2}(x1, x2)dx_1 dx_2 = 1$$

Joint cdf, $F_{X_1,X_2}(x_1, x_2)$:

-

$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_{X1,X2}(t_1, t_2) dt_1 dt_2$$

-

$$\frac{\partial^2}{\partial x_1 \partial x_2} F_{X_1,X_2}(x_1, x_2) = f_{X_1,X_2}(x_1, x_2)$$

1.4.1.1   Dependence and Independence

If and only if X_1 and X_2 are independent, then

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

Expectation and variance-covariance matrix:

$$E \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} == \begin{pmatrix} E(X_1) \\ E(X_2) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\Sigma = Cov \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{bmatrix}$$

Covariance is $Cov(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$, so:

$$Cov(X_1, X_1) = E(X_1^2) - E(X_1)^2 = Var(X_1)$$

Note that the variance-covariance matrix is symetrial because on the left-to-right diagonal it has variances and on the right-to-left diagonal they're the same.

$$Cov(X1, X2) = E\{(X_1 - E(X_1))(X_2 - E(X_2))\}$$

If the X_1 and X_2 are independent, $Cov(X_1, X_2) = 0$. But this in *not* a biconditional.

Proof.

$$E(X_1, X_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X_1 X_2 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) x_2 f_{X_2}(x_2) dx_1 dx_2$$

$$= \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2$$

$$= E(X_1) E(X_2) = Cov(X_1, X_2) = E(X_1, X_2) - E(X_1) E(X_2) = 0$$

Conditional distributions

$$f_{X_1 | X_2 = x_2}(x_1, x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

Marginal distributions:

Individual distribution of x_1:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_2$$

□

Example 1

$X_1 \sim Unif(0, 1)$, $X_2 \sim Unif(0, 1)$, X_1 and X_2 are independent.

$$f_{X_1}(x_1) = \frac{1}{1 - 0} = 1$$

$$f_{X_2}(x_2) = \frac{1}{1 - 0} = 1$$

(a) Find the joint pdf of $\binom{X_1}{X_2}$

$$f_{X_1,X_2}(x_1, x_2) = 1 \times 1 = 1$$

(b) Find the joint cdf:

$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \leq x_1 \cap X_2 \leq x_2) = P(X_1 \leq x_1)P(X_2 \leq x_2) = x_1 x_2$$

In general: $F_{X_1,X_2}(1,1) = 1$, $F_{X_1,X_2}(0,0) = 0$, $F_{X_1,X_2}(\infty, \infty) = 1$, $F_{X_1,X_2}(-\infty, -\infty) = 0$.

(c) Find $E\binom{X_1}{X_2}$:

$$E\binom{X_1}{X_2} = \binom{E(X_1)}{E(X_2)} = \binom{0.5}{0.5}$$

(d) Find $Cov(X_1, X_2) = \Sigma$

$$\Sigma = Cov\binom{X_1}{X_2} = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) \end{bmatrix} = \begin{bmatrix} 1/12 & 0 \\ 0 & 1/12 \end{bmatrix}$$

Zeros from right-to-left diagonals because X_1 and X_2 are independent.

Example 2

Joint pdf $f_{X_1,X_2} = c(x_1^2 + x_2^2)$, where c is the normalization constant and $x_1, x_2 \in [0, 1]$

(a) Find c

$$1 = \int_0^1 \int_0^1 c(x_1^2 + x_2^2)dx_1 dx_2$$

$$= c \int_0^1 \int_0^1 (x_1^2 + x_2^2)dx_1 dx_2$$

...

$$= 2/3(c), \quad x_1, x_2 \in [0, 1]$$

So $c = 3/2$.

$$f_{X_1 X_2} = \frac{3}{2}(x_1^2 + x_2^2) \quad x_1, x_2 \in [0, 1]$$

(b) Find joint cdf:

$$F_{X_1 X_2}(x_1, x_2) = \int_0^{x_1} \int_0^{x_2} \frac{3}{2}(t_1^2 + t_2^2) dt_1 dt_2 = \frac{x_1^3 x_2 + x_1 x_2^3}{2}, \quad x_1, x_2 \in [0, 1]$$

(c) Find the marginal pdf and cdf of x_1:

$$f_{X_1}(x_1) = \int_0^1 \frac{3}{2}(x_1^2 + x_2^2) dx_2$$

$$= \frac{3}{2}\left(x_1^2 + \frac{1}{3}\right), \quad x_1, x_2 \in [0, 1]$$

$$F_{X_1}(x_1) = \int_0^{x_1} \left(\frac{3}{2}x_1^2 + \frac{1}{2}\right) dx_1 = \frac{1}{2}x_1^3 + \frac{1}{2}x_1, \quad x_1, x_2 \in [0, 1]$$

(d) $Cov(X_1, X_2) = E(X_1, X_2) - E(X_1)E(X_2)$

$$E(X_1, X_2) = \int_0^1 \int_0^1 x_1 x_2 f_{X_1 X_2}(x_1, x_2) dx_1 dx_2 = \int_0^1 \int_0^1 x_1 x_2 \frac{3}{2}(x_1, x_2) dx_1 dx_2 =$$

$$E(X_1) = \int_0^1 x_1 f_{X_1}(x_1) dx_1 = \int_0^1 \left(\frac{3}{2}x_1^3 + \frac{1}{2}x_1\right) dx_1 = \frac{5}{8}$$

$$E(X_2) = E(X_1) = \frac{5}{8}$$

$$Cov(X_1, X_2) = \frac{3}{8} - \left(\frac{5}{8}\right)^2 = \frac{-1}{64}$$

X_1 and X_2 could be independent if $f_{X_1 X_2} = x_1 x_2$ because they can be products of their own parts. $f_{X_1 X_2} = x_1 x_2 + x_1$ could also be independent because you can factor it into $f_{X_1 X_2} = x_1(x_2 + 1)$.

If it's $f_{X_1 X_2} = x_1 x_2 + x_1 + x_2$ then you can't factor them, so they're not independent.

1.4.2 Coorelation

This defines the strength of linear relationship between X_1 and X_2:

$$r = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}} = \frac{E(X_1, X_2) - E(X_1)E(X_2)}{\sqrt{E(X_1^2) - E(X_1)^2}\sqrt{E(X_2^2) - E(X_2)^2}}$$

Correlation can only be between two variables. With more there can only be pairwise correlation.

$$X_1 \text{and} X_2 \text{ are independent} \implies r = 0$$

$$X_1 \text{and} X_2 \text{ are independent and jointly normal} \iff r = 0$$

$$\Sigma_x = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$X \sim N(\mu_x, \Sigma_x)$

Since there are 5 parameters, sometimes we write:

$$X \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r)$$

$$f_{X_1 X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x_1^2}{2}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x_2^2}{2}} = \frac{1}{2\pi} e^{\frac{-x_1^2 - x_2^2}{2}}$$

The *affine transform* is $Y = AX + B$ where A is the deterministic 2x2 matrix and B is a 2-dimensional vector.

$$E(Y) = E(AX + B) = AE(X) + B = A\mu_x + B$$

$$Cov(Y) = E((Y - E(Y))(Y - E(Y))^T) = E((AX + B - (A\mu_x + B))(AX + B - (A\mu_x + B)^T))$$

B cancels, so:

$$= E(A(X - \mu_x)(X - \mu_x)^T A^T) = AE((x - \mu_x)(x - \mu_x)^T)A^T$$

In our case, $B = ...$:

$$E(Y) = A\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

1.4.3   Joint distributions, marginals, conditionals

Conditional distributions will be on final. Matrix multiplication.

1.4.4   Covariance and Correlation

1.4.5   Multivariance Normal

These will come up on the final in Reuben's part.

IID, independently identically distributed.

X_1, X_2 IID $N(0,1)$. $X_1 \sim N(0,1)$ and $X_2 \sim N(0,1)$.

Consider $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ bivariate normal distrib.

$$E(X) = \begin{pmatrix} E(X_1) \\ E(X_2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Chapter 2

Stochastic Signals and Systems Applied to EE

This chapter covers how to use stochastic signals and systems in practical applications related to electrical engineering.

Important applications include *telecommunications*, in which noise and interference create uncertainty and hamper the ability to be reliable; *statistical signal processing*, including speech and image compression, reconstruction, transmission, recognition, as well as topics like *machine learning*.

For example, even electrical systems behave in 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 when doing calculations like Ohm's Law.

In physics and biology, quantum mechanics, molecular, electro-diffusive, and fractional dynamics all involve statistics, as does mathematical finance and economics.

In all of these applications, construction of the event-space definition is the most challenging task, both in industry and in academia.

2.1   Probability Review

2.1.1   Review of Random Variables

Ω is the set of all possible outcomes. Where an event is a subset of Ω.

1. $0 \leq P(A) \leq 1$ for any event A

2. $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$ (A and B are mutually exclusive):

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

3. $P(\Omega) = 1$, $P(\emptyset) = 0$

P is a function that maps any event A to $[0, 1]$. You could think of it as measuring how big the event (a set) is. For instance:

$$A \subset B \implies P(A) \leq P(B)$$

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

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

For **mutually exclusive events**, $A \cap B = \emptyset$. Meaning that both cannot occur, or that the sets are disjoint.

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

For **independent events**, meaning that the fact that event A occurs has no affect on if event B occurs:

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

Mutually exclusive events are *not* necessarily independent.

If A and B are independent and B and C are independent, A and B are *not* necessarily independent.

If $A \subset B$, then A and B are not independent because $P(A \cap B) = P(A) \neq P(A)P(B)$.

If $A \subset B$, then A and B are not mutually exclusive because $P(A \cup B) = P(B) \neq P(A) + P(B)$

If $A \cap B = \emptyset$, then A and B are *not* necessarily independent. If $P(A) \neq 0$ and $P(B) \neq 0$, then for them to be independent $P(A \cap B) \neq 0$, but for them to be mutually exclusive, $P(A \cap B) = 0$ must be true.

Conditional probability states:

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

Additionally,

$$P(A|B) \leq \frac{P(A)}{P(B)}$$

Because

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) + P(B) - P(A \cup B)}{P(B)} = \frac{P(A)}{P(B)} + 1 - \frac{P(A \cup B)}{P(B)}$$

And we know $P(A \cup B) \geq P(B)$, so $1 - \frac{P(A \cup B)}{P(B)} \leq 0$.

If A and B are independent, then $P(A|B) = P(A)$.

If $A = B$, then $P(A|B) = P(A|A) = 1$.

Total Probability states for mutually exclusive events C_1, C_2, ..., C_n that form a complete partion of Ω, $C_1 \cup C_2 \cup \cdots \cup C_n = \Omega$, then for any event A:

$$\begin{aligned} P(A) &= P(A \cap C_1) + P(A \cap C_2) + \cdots + P(A \cap C_n) \\ &= P(A|C_1)P(C_1) + P(A|C_2)P(C_2) + \cdots + P(A|C_n)P(C_n) \\ &= \sum_{i=1}^{n} P(A|C_i)P(C_i) \end{aligned}$$

Example

Let $\{ C_i \}_{i=1}^{n} = \Omega$ be a partition of sample space Ω and A and B be two events. If (a) A and B are conditionally independent given C_i and (b) B is independent of all C_i, prove that A and B are independent.

From (a) we know $P(A \cap B|C_i) = P(A|C_i)P(B|C_i)$. From (b) we know $P(B|C_i) = P(B)$.

$$\begin{aligned} P(A \cap B) &= \sum_{i=1}^{n} P(A \cap B|C_i)P(C_i) \\ &= \sum_{i=1}^{n} P(A|C_i)P(B|C_i)P(C_i) && \text{From (a)} \\ &= \sum_{i=1}^{n} P(A|C_i)P(B)P(C_i) && \text{From (b)} \\ &= P(B)\sum_{i=1}^{n} P(A|C_i)P(C_i) \\ &= P(B)P(A) && \text{By total probability} \end{aligned}$$

CDF: $F_X(x) = P(X \le x)$, $P(a < X \le b) = F_X(b) - F_X(a)$

$F_{X,Y}(x,y) = F_X(x)F_Y(y)$

PMF/PDF: $f_X(x) = P(X = x_i)$

$$F_X(x) = \int_{-\infty}^{\infty} f_X(s)ds$$

If $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, then X and Y are **independent**.

For a continuous random variable, $P(X = a) = 0$

Expected value/mean:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$$

For any function g, (???)

$$E(X) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

$E(aX + bY) = aE(X) + bE(Y)$

Variance:

$$\text{Var}(X) = E(X - \mu^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)dx = E(X^2) - E(X)^2$$

$\text{Var}(aX + b) = a^2$

Covariance:

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y)f_{X,Y}(x, y)dxdy$$

$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$

If $\text{Cov}(X, Y) = 0$, then X and Y are **uncorrelated**.

Note $f_{X|Y}(x|y) = f_X(x)$ and $F_{X|Y}(x|y) = F_X(x)$

2.1.1.1 Uniform Random Variables

2.1.1.2 Exponential Random Variables

Exponential distributions are the only memoryless continuous random variable.

Weird things:

1. If the reliability of a car was exponentially distributed, then a used car is just as good as a brand new car

2. Suppose a bus arrives on average every 10 minutes, $\lambda = 1/10$. If you have already waited 9 minutes, the probability of the bus arriving in the next minute is the same as if you have not waited for the 9 minutes.

Proof.

$$
\begin{aligned}
P(X > x + \Delta | X > \Delta) &= \frac{P(X > x + \Delta \cap X > \Delta)}{P(X > \Delta)} \\
&= \frac{P(X > x + \Delta)}{P(X > \Delta)} \\
&= \frac{e^{-\lambda(x+\Delta)}}{e^{-\lambda\Delta}} \\
&= e^{-\lambda\Delta} \\
&= P(X > x)
\end{aligned}
$$

$\square$

2.1.1.3   Gaussian Random Variables

Example

If $Y = aX + b$ and $X \sim N(\mu, \sigma^2)$, compute $f_Y(y)$.

$$\mu_y = E(Y) = E(aX + b) = aE(x) + bE(1) = a\mu + b$$

$$\sigma_y^2 = \text{Var}(y) = a^2\sigma^2$$

2.1.1.4   Binomial Random Variables

2.2   Transformations of Random Variables

Can move from any distribution to any other distribution through a series of transformations. These transformations have many applications including generating random variables, analysis of transformers, superposition of source signals from different sensors, and analysis.
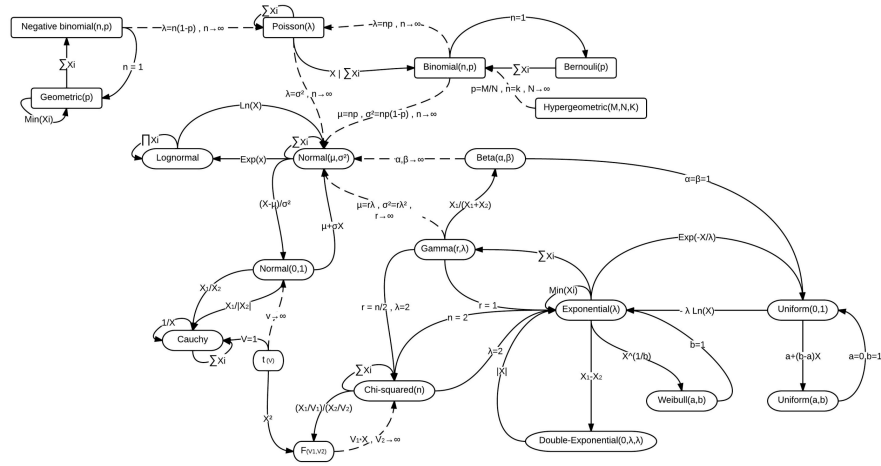
Figure 2.1: Transformation Relationships Between Distributions

There are three main methods of transformation:

1. Univariate Transform
2. Multivariate Transform
3. Probability Integral Transform

2.2.1 Univariate Transform

This is a transformation for one variable. Suppose we have a random variable X and we want to get the probability distribution of some function of X, we can do this.

ie. We have $X \sim f_X(x)$ and we want to get $f_Y(y)$ where $Y = g(X)$.

To do this we have to make the assumption that $g(X)$ is a one-to-one/injective function, that is $y = g(x)$ has only one solution (passes the horizontal line test).

Making that assumption we can solve for the cdf of Y like so:

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(g(X) \le y) \\
&= \int_{\{x \mid g(x) \le y\}} f_X(x) dx \\
&= \int_{-\infty}^{g^{-1}(y)} f_X(x) dx
\end{aligned}
$$

$$
\boxed{F_Y(y) = \int_{-\infty}^{g^{-1}(y)} f_X(x) dx} \tag{2.1}
$$

Univariate Transform cdf

Then determine the pdf by taking the derivative with respect to y:

$$
\begin{aligned}
f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\
&= \frac{\partial}{\partial y} \int_{-\infty}^{g^{-1}(y)} f_X(x) dx \\
&= \left[f_X\left(g^{-1}(y)\right) \frac{\partial}{\partial y} g^{-1}(y) \right] - \left[f(-\infty) \frac{\partial}{\partial y} (-\infty) \right]^{\,0} \qquad \text{By Leibniz integral rule} \\
&= f_X\left(g^{-1}(y)\right) \left| \frac{g^{-1}(y)}{dy} \right| \qquad\qquad\qquad\qquad \text{Absolute because } f_Y(y) \ge 0
\end{aligned}
$$

$$
\boxed{f_Y(y) = f_X\left(g^{-1}(y)\right) \left| \frac{g^{-1}(y)}{dy} \right|} \tag{2.2}
$$

Univariate Transform pdf

Example

Suppose we have a noiseless non-Inverting amplifier. It takes in a signal x and outputs the signal $y = ax$.

If x comes from Gaussian noise source $X \sim N(\mu, \sigma^2)$, what is the distribution of the output signal $f_Y(y)$?

Recall we make the assumption that $y = g(x)$, which means $g^{-1}(y) = x$, so:

$$g^{-1}(y) = x = \frac{y}{a}$$

Using (2.2), we know:

$$
\begin{aligned}
f_Y(y) &= f_X\left(g^{-1}(y)\right)\left|\frac{g^{-1}(y)}{dy}\right| \\
&= f_X\left(\frac{y}{a}\right)\left|\frac{\frac{y}{a}}{dy}\right| \\
&= \frac{1}{a}f_X\left(\frac{y}{a}\right)
\end{aligned}
$$

Plugging in $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ and simplifying, you get this relation:

$$f_Y(y) = N(a\mu, a^2, \sigma^2)$$

Now suppose $Y = g(X)$ is not injective, so we can't assume $x = g^{-1}(x)$, then how would we get $f_Y(y)$ given $f_X(x)$? Well, we could directly evaluate the following:

$$f_Y(y) = \frac{\partial}{\partial y}\left(\int_{\{x|g(x)\leq y\}} f_X(x)dx\right)$$

But this is hard because we don't know what $\{x|g(x) \leq y\}$ is. So if $g(x)$ is not injective, then we can denote the n roots of $y = g(x)$ by $\{X_i\}_{i=1}^n$. Then use the following, where $x_i = x_i(y)$:

$$\boxed{f_Y(y) = \sum_{i=1}^n f_X(x_i)\left|\frac{\partial x_i}{\partial y}\right|} \tag{2.3}$$

Univariate transformation relation

Example

A non-inverting amplifier takes an input signal x and outputs the signal $y = ax^2$.

If x comes from Gaussian noise $X \sim N(\mu, \sigma^2)$, what is the distribution of output signal $f_Y(y)$?

First determine the routes of $y = g(x) = ax^2$.

$$x_1 = \sqrt{\frac{y}{a}} \qquad\qquad \frac{\partial x_1}{\partial y} = \frac{1}{2y}\sqrt{\frac{y}{a}}$$

$$x_2 = -\sqrt{\frac{y}{a}} \qquad\qquad \frac{\partial x_2}{\partial y} = -\frac{1}{2y}\sqrt{\frac{y}{a}}$$

Now using (2.3):

$$
\begin{aligned}
f_Y(y) &= \sum_{i=1}^{n} f_X(x_i) \left| \frac{\partial x_i}{\partial y} \right| \\
&= f_X(x_1) \left| \frac{\partial x_1}{\partial y} \right| + f_X(x_2) \left| \frac{\partial x_2}{\partial y} \right| \\
&= f_X\left(\sqrt{\frac{y}{a}}\right)\left[\frac{1}{2y}\sqrt{\frac{y}{a}}\right] + f_X\left(-\sqrt{\frac{y}{a}}\right)\left[\frac{1}{2y}\sqrt{\frac{y}{a}}\right] \\
&= \frac{1}{2\sqrt{ya}}\left[f_X\left(\sqrt{\frac{y}{a}}\right) + f_X\left(-\sqrt{\frac{y}{a}}\right)\right]
\end{aligned}
$$

2.2.2 Multivariate Transform

Now given $V = g(X,Y)$, $h(X,Y)$, and $f_{XY}(x,y)$, we're going to learn how to get the joint pdf $f_{VW}(v,w)$.

Theoretically we could do something like:

$$f_{VW}(v,w) = \frac{\partial}{\partial v}\frac{\partial}{\partial w} \iint\limits_{\{(x,y) \ | \ g(x,y)\leq v, \ h(x,y)\leq w\}} f_{XY}(x,y)\,dxdy$$

But that's just ridiculous. It's impossible to visualize and begin defining event space here, so let's look at how to evaluate $f_{VW}(v,w)$ without needing

explicit evaluation of multivariate integral limits:

$$f_{VM}(v,w) = \sum_{i=1}^{n} f_{XY}(x_i, y_i)\left|\tilde{\mathbf{J}}_i\right| \quad \text{where } \left|\tilde{\mathbf{J}}_i\right| = \left|\frac{\partial\phi_i}{\partial v_i}\frac{\partial\psi_i}{\partial w_i} - \frac{\partial\phi_i}{\partial w_i}\frac{\partial\psi_i}{\partial v_i}\right| \quad (2.4)$$

Multivariate transformation relation

Here the inverse-mappings of (v, w) are given by $\{x_i = \psi_i(v, w), \ y_i = \phi_i(v, w)\}_{i=1}^{n}$, and $\left|\tilde{\mathbf{J}}\right|$ is the magnitude of the *Jacobian*, which is equal to $A(\mathbb{R})/A(\mathbb{N})$.

Example

$V = g(X, Y) = \sqrt{X^2 + Y^2}$ and $W = h(X, Y) = Y/X$.

If $X \sim N(0, \sigma^2)$ and $Y \sim N(0, \sigma^2)$ are independent distributions, evaluate (a) $f_{VW}(v, w)$, and (b) show $f_{VW}(v, w)$ is independent.

Solving the system:

$$V = \sqrt{X^2 + Y^2}$$
$$W = Y/X$$

We get solution 1:

$$x_1 = \phi_1(V, W) = \frac{V}{\sqrt{1 + W^2}}$$
$$y_1 = \psi_1(V, W) = \frac{WV}{\sqrt{1 + W^2}}$$

And solution 2:

$$x_2 = \phi_2(V, W) = \frac{-V}{\sqrt{1 + W^2}}$$
$$y_2 = \psi_2(V, W) = \frac{-WV}{\sqrt{1 + W^2}}$$

$$\left| \tilde{\mathbf{J}}_1 \right| = \left| \frac{\partial \phi_1}{\partial v} \frac{\partial \psi_1}{\partial w} - \frac{\partial \phi_1}{\partial w} \frac{\partial \psi_1}{\partial v} \right|$$

$$= \left| \frac{\partial}{\partial v} \left(\frac{V}{\sqrt{1+W^2}} \right) \frac{\partial}{\partial w} \left(\frac{WV}{\sqrt{1+W^2}} \right) - \frac{\partial}{\partial w} \left(\frac{V}{\sqrt{1+W^2}} \right) \frac{\partial}{\partial v} \left(\frac{WV}{\sqrt{1+W^2}} \right) \right|$$

$$= \left| \left(\frac{1}{\sqrt{1+W^2}} \right) \left(\frac{V}{(1+W^2)^{3/2}} \right) - \left(\frac{-WV}{(1+W^2)^{3/2}} \right) \left(\frac{W}{\sqrt{1+W^2}} \right) \right|$$

??? In his notes in `Lecture07hw.pdf` page 3, he says:

$$\frac{\partial y_1}{\partial v} = \frac{\partial}{\partial v} \left(\frac{wv}{1+w^2} \right) = \frac{w}{1+w^2} = 0?$$

Example

Let $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$ be the random variables associated with two independent sensors that provide an estimate of the position of a target.

A non-symmetric differential amplifier merges the signals, producing an estimate of the target's position. The noise of the estimate is given by $Z = aX - Y$ with $a \geq 0$.

(a) Compute the pdf of Z

(b) Evaluate μ_z and σ_z^2 using $f_Z(z)$

(c) What is the optimal a for zero average noise? What's the noise variance of this a?

(a)

2.2.3 Probability Integral Transform

You can use this to test if your dataset $\{x_i\}_{i=1}^{n}$ resulted from a certain distribution $f_X(x)$.

Given a cdf $F_X(x)$ then the random variable $Y = F_X(X)$ has a uniform distribution.

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(F_X(X) \leq y) \\
&= P(X \leq F_X^{-1}(y)) \\
&= F_X(F_X^{-1}(y)) \\
&= y
\end{aligned}
$$

Given a continuous random variable $Y = F_X(X)$, you can use the *probability integral transform*, you cannot with discrete random variable samples.

$$\boxed{F_X(X) \sim \text{Unif}(0,1)} \tag{2.5}$$

Probability Integral Transform

2.3 Generating Distributions

We can use computers to generate random numbers with a specified probability density. We do this to model real world parameters that vary with these probabilities.

Stochastic means that you don't actually know exactly what a parameter's value is, you know with a certain probability what its range is.

2.3.1 Inverse Transform Algorithm

2.3.1.1 Discrete

The inverse transform algorithm is useful for any discrete distribution. Given a uniform random number generator, how can we simulate a discrete distribution $P_X(X = x_i) = p_i$?

1. Generate random variable $u \sim U(0, 1)$

2. Set $x = \begin{cases} x_1 & u \in [0, p_1] \\ x_2 & u \in (p_1, p_1 + p_2) \\ \vdots & \vdots \\ x_n & u \in (\sum_{i=1}^{n-1} p_i, 1] \end{cases}$

Let's say:
$$x = \begin{cases} 1 & 20\% \text{ of the time} \\ 2 & 10\% \text{ of the time} \\ 3 & 70\% \text{ of the time} \end{cases}$$

Then you generate a uniform and if it falls between $[0, 20]$, you assign the value 1. If the uniform falls between $(20, 30)$, you assign value 2. Then for $\in (30, 70]$, you assign 3. You get the idea.

2.3.1.2   Continuous

In the case that the distribution you want, say $F_X(x)$ is continuous, then it'll look like this:
$$F_X(x) = P(X \le x) = p$$
It takes x as an input and outputs p. If you take the inverse of that function so:
$$x = F_X^{-1}(p)$$

Then the input will be p, which is actually uniform so we'll call it u, and given uniform inputs will provide outputs x in the desired distribution.

Note: You cannot use the inverse transform method to generate a Gaussian because there's no good CDF to take the inverse of.

The opposite of this is that if you want to generate a uniform, then you generate samples of a distribution, then you feed those values into that distribution's CDF and you'll get out a uniform distribution.

2.3.2   Polar Algorithm

You can construct anything from this method, it's particularly useful for generating a Gaussian which we can't do with the inverse method.

1. Generate $u_1 \sim U(0,1)$, and $u_2 \sim U(0,1)$

2. Set $d = -2\ln u_1$ and set $\theta = 2\pi u_2$

3. Set $x = \sqrt{d}\cos\theta$ and $y = \sqrt{d}\sin\theta$

And by doing this, $X \sim N(0,1)$ and $Y \sim N(0,1)$ are independent normal random variables.

To then construct a non-standard normal distribution, use the linear properties and set:
$$Z = \sigma X + \mu \quad \text{where } X \sim N(0,1)$$

2.3.3 Compositional Method

If it's challenging to create a distribution using the inverse transform method, you can try the compositional method.

The idea here is to decompose your desired distribution $F_X(x)$ from a sum of weighted distributions F_{X_i}. This is good if it's easy to generate samples from the F_{X_i} distributions.

$$F_X(x) = \sum_{i=1}^{n} p_i F_{X_i}(x) \quad \text{where} \sum_{i=1}^{n} p_1 = 1 \quad p_i \geq 0$$

Then we can use the compositional method to generate samples from $F_X(x)$

1. Generate random variable I which has a discrete distribution $P(I = i) = p_i$, for $i = 1, 2, \ldots, n$.

2. Generate random variable Y_I with distribution F_{X_I}

3. Set $X = Y_I$

2.3.4 Acceptance-Rejection Method

Propose value, see if it gets accepted. Often use exponential because it goes from $[0,\infty]$ and looks like a symmetrical side of normal.

Sampling something constrained between values use uniform, sampling something that's positive use exponential, sampling something that uses entire real line use normal.

2.4 Random Processes

2.4.1 Independent and Identically Distributed (IID)

With IID distributions, each random variable has the same distribution (possible outcome) and all outcomes are mutually independent.

A random process is IID if:

1. It has no memory, so past results tell us nothing about future results

$$P(X[k] = x_k | X[1], \ldots, X[k-1]) = P(X[k] = x_k) = f_{X_k}(x_k; k)$$

2. Identical/time-invariant distribution

$$P(X[k] = x_k) = f_{X_k}(x_k; k) = f(x_k)$$

And IID process is actually a special case of a Markov chain, which we get to in section 2.4.2.

2.4.2 Markov Chain

Most sources have some sort of correlated random process, and are not pure IID. Markov Chains are a very important type of dependent stochastic process, and almost all engineering, economics, and social sciences model is Markov.

Whereas IID processes have no memory:

$$\boxed{P(X[k] \mid X[k-1], X[k-1], \ldots) = P(X[k])} \tag{2.6}$$

IID Process

Markov chains do have some sort of memory:

$$\boxed{P(X[k] \mid X[k-1], X[k-1], \ldots) = P(X[k] \mid X[k-1]) \forall k} \tag{2.7}$$

First-order Markov chain

They are characterized by three ingredients:

1. **State space** – $S = \{x_1, \ldots, x_m\}$

2. **Initial state distribution** – $\pi[0]$ (M dimensional column vector)

   $\pi[0] = [P(X[0] = x_1), P(X[0] = x_2), \ldots, P(X[0] = x_M)]'$

   Clearly, $\pi'[0]\mathbf{1} = 1$ – ??? WHAT THE FUCK DOES THIS MEAN

3. Transition probability matrix: $A = (a_{ij})$ which is an $M \times M$ matrix

$$\sum_{j=1}^{M} a_{ij} = 1 \implies A_{M \times M}\mathbf{1}_{M \times 1} = \mathbf{1}_{M \times 1}$$

   In this matrix A, each row sums to 1, and all elements are non-negative.

Example

Consider a two-state Markov chain $(M = 2)$ with $\pi[0] = [0.3, 0.7]'$

$A = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$
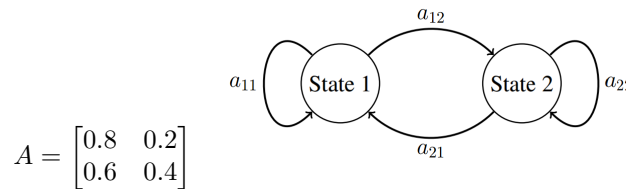
Remarks about this Markov chain

- $a_{ij} = P(X[x+1] = j \mid X[k] = i)$, denotes the transition probability from state i to state j. a_{ii} denotes the probability of remaining in state i.

- A always has one eigenvalue at $\lambda = 1$ corresponding to eigenvector $\mathbf{1}$ because $A\mathbf{1} = \mathbf{1}$

- A^k is always a stochastic matrix for any integer $k > 0$ — meaning this is a first-order *homogeneous* Markov Chain because A is not a function of time (???)

- Markov property means that there is *conditional independence*, in that given $X[k]$, $X[k+1]$ is independent of the past (???)

2.4.2.1 Stationary Distribution

How can we compute the **steady-state probability** (stationary distribution), $\pi[\infty]$ of a Markov chain?

If we calculate $\pi[\infty]$ then we can model any source by a Markov chain, estimate the transition probabilities, and compute how much info it has.

We do this using the **Chapman-Kolomorov Equation**. Given $\pi[0]$ and A, the state probability vector at time k is given by:

$$\pi[k] = [P(X[k] = x_1), \ldots, P(X[k] = x_M)]'$$

This can be computed by:

$$\boxed{\pi[k+1] = A'\pi[k] = (A')^{k+1}\pi[0]}$$
(2.8)

Chapman-Kolmogorov Equation

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

$0 \le P(A \cup B) \le 1$, then the minimum bound is where $P(A \cup B) = 1$ and:

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

Example 2

Prove:

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

—

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 $P(A_1 \cap A_2) \ge 0$, then

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

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

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

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

$$P(\cup_{i=1}^{k+1} A_i) = P(\cup_{i=1}^{k} A_i \cup A_{k+i}) \le P(\cup_{i=1}^{k} A_i) + P(A_{k+1})$$

$$\le \sum_{i=1}^{k} P(A_i) + P(A_{k+1}) = \sum_{i=1}^{k+1} P(A_i)$$

$\square$

Problem 1.1

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

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

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

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

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

(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...

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

69

(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
    $$F_X(B) = 1 = P(X \leq B) \text{ and } F_X(a) = 0 = P(X \leq a)$$

    (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 $\Phi(c) = \texttt{pnorm(c)}$.

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

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 $c = \texttt{qnorm(0.75)} = 0.674$. 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

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

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

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?

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

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

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

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

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

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

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

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

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

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)?

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

Standardize, subtract the mean.

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

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

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

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 $P(X < 10) \le 0.1$, we need $\mu \ge 10.128$.

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:

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

$$\frac{10 - \mu}{0.1\mu} = -1.216(???) \to \mu \ge 11.47$$

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.

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

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

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

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

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

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

79

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

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

$$P(X > 1)$$

E   In-Class Tutorial

E.1   Binomial Random Variable

Models the number of successis in a sample of size n with a replacement from a population of size M. It also works in M is very, very large, $n << M$.

(a) Compute the *cdf* of $f_X(x)$

In the continuous cade, you can bring the discrete domain to the continuous domain using the dirac delta function.

$$F_X(x) = \int_{-\infty}^{x} f_X(t) \sum_{i=1}^{n} \delta(t - x_i)dt$$

(b) Compute the mean

E.2 Resistor Selection

Suppose we choose resistor R from a batch of resistors with $\mu = 1000\Omega$ and $\sigma = 200\Omega$.

To get these parameters, you can get a sample like so:

Sample mean: $\hat{\mu} = 1/N \sum_{i=1}^{N} R_i$

Sample variance: $\hat{var}^2 = 1/N \sum_{i=1}^{N} (R_i - \mu)^2$

By the weak law of large numbers, if N is sufficiently large, then $\hat{\mu} = \mu$ and $\sigma^2 = \hat{\sigma}^2$.

(a) With no further information, what is a reasonable distribution of R?

—

If you're given a mean and a variance, it's natural to assume that it's a normal distribution, but there is no actual reason to assume that. You should perform other statistical tests to confirm.

But anyway, let's say $R \sim N(\mu, \sigma^2)$

(b)
$$P(900 < R \le 1100) = \int_{900}^{1100} f_R(r)dr = erf(0.5) - erf(-0.5)$$

(???)

E.3 Failure Estimate

A company produces one defective chip for every five good chips. The defective chips have a time of failure X that obeys the cumulative distribution function.

$$F_{X|Y}(x|Y=0) = (1 - e^{-x/2})u(x)$$

—

Given information: 1 defective chip per 5 chips. $Y \in \{0,1\}$ is a binary random variable that specifies if the chip is defective or not. These, defective and non-defective, are two mutually exclusive events. $X \in (0, \infty)$ is the lifetime of the random variable.

Conditional distributions:

$$F_{X|Y}(x|y=0) = (1 - e^{-x^2/2})u(x)$$

$$F_{X|Y}(x|y=1) = (1 - e^{-x^2/10})u(x)$$

These are conditional Poisson distributions with of average rate of failure of 2 and 10.

(a) Compute cdf. —

Events $A_1 = \{Y = 0\}$ and $A_2 = \{Y = 2\}$. Then $P(A_1 \cap A_2) = 0, P(A_1 \cup A_2) = 1$, so $A_1 \cup A_2 = \Omega$.

Define another event $B = \{$any event... (???) $\}$.

$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2)$.

$B = B \cap \Omega = (B \cap A_1) \cup (B \cap A_2)$

By definition, the intersection operator $B \cap A_i \subset A_i$.

Since $(B \cap A_1) \cup (B \cap A_2) = \emptyset$, these events are mutually exclusive.

From axiom of probibility and cond. probability,

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

$= P(B \cap A_1) + P(B \cap A_2) = P(B|A_1)P(A_1) + P(B|A-2)P(A_2)$.

Putting it all together:

$$P(A_1) = P(Y = 0) = 1/6$$

$$P(A_2) = 1 - P(A_1) = 5/6$$

(b) Probability of failure around 6 months $P(x \le 6) = 0.5...(???)$

E.4 Conditional Poisson

Let X be a Poisson random variable with rate parameter $\lambda > 0$.

$$f_X(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \quad x \in \{0, 1, 2...\}$$

Define the event $Y = \{x \text{ is even}\} = \{0,2,4...\}$

(a) Compute the conditional pdf

—

Use Baye's Rule, for conditional distribution:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

$f_Y(y) = P(\{x \text{ is even}\})$

To make things easier, define $z = \{x \text{ is odd}\}$

$f_Z(z) = P(\{x \text{ is odd}\})$

Property of factorial sums and $f_Y(y) + f_Z(z) = 1$ ... (???)

$\rightarrow f_Y(y) = \frac{1}{2}(1 + e^{-2x})$

Now we need an expression for the joint distribution $f_{XY}(x, y)$

$f_{XY}(x, y) = P(X = x, X \text{ is even})$

$X = x \cap X \text{ is even} = \emptyset$ if x is odd.

$X = x \cap X \text{ is even} = \emptyset$ if x is odd. ...(???)

Putting it together:

...

84

F   Tutorial 4 (Yang)

Example

$F_X(x) = 1 - \frac{1}{x^2}$, $x > 1$

(a) Find pdf of $y = x^2 \rightarrow g(x) = x^2$

CDF of given:

$$F_X(x) = 1 - \frac{1}{x^2}$$

Since $x > 0$:

$$F_Y(y) = P(Y \leq y) = P(x^2 \leq y)$$

$$F_Y(y)$$

not paying attention...

Hint for assignment 2, Q1

(a)

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \Sigma_x \right)$$

$$\begin{pmatrix} v \\ w \end{pmatrix} = f\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} v \\ w \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix}$$

$$sin\begin{pmatrix} x \\ y \end{pmatrix} \sim N \implies \begin{pmatrix} v \\ w \end{pmatrix} \sim N(\hat{u}, \Sigma)$$

$$Z = 2X + 3Y, X, Y \sim N(Qu, \delta)$$

Q2

Q3

Generate true density directly from expressions in (a) and (b).

...

Inverse transform

Random variable X with pdf $f_X(x)$ and $F_X(x)$ is invertible.

If we want to generate N values of X:

Algorithm

Step 0: compute the cdf $F_X(x)$.

Repeat the following N times:

    Step 1: Generate a uniform RV $u = runif(I)$

    Step 2: Apply $F_X^{-1}(u)$ and record result as x

Compute $F_X^{-1}(u)$ by $F_X(x) = u$ and solve x

MATLAB pseudocode:

```
u = rand(1e5, 1);
x = F_x^{-1}(u);
histogram(x);
compare
```

Example

Generate 10,000 values from: $f_X(x) = \int^x (f_X(x))dt = ln(x)$

To get inverse pdf: $ln(x) = u \rightarrow x = e^u$, $F_x^{-1}(u) = e^u$

Inverse transform for discrete random

RV X discrete, $f_X(x) = P(X = x)$ — pmf

Algorithm

Step 1: Generate a Unif(0,1) $= u$

Step 2:

$$Set\ x = \begin{cases} x_1 & 0 \le u \le p_1 \\ x_2 & p_1 \le u \le p_1 + p_2 \\ ... & \\ x_n & \sum_{i=1}^{n-1} p_z \le u \le 1 \end{cases}$$

Example

10,000 values

$X - P(X = x)$

$1 - \frac{1}{3}$

$2 - \frac{1}{6}$

$3 - \frac{1}{4}$

$4 - \frac{1}{4}$

Generating exponential distribution in MATLAB:

```
mu = 1;
u = rand(1e5,1);
X = expinv(1e6,1,me);
numbins = 100;
hist(X, numbins, 'g'); hold on;


h = hist(X, numbins, 'g');


v = 0:0.01:10;
Y = exp(-v);
plot(v, Y*h(1), 'r', 'LineWidth', 2);
```

Compositional Method

Can use when $F_X(x) = \sum_{i=1}^{n} P_i F_{x_i}(x)$, $\sum_{i=1}^{n} P_i = 1$, $P_i \geq 0$

Where P_i is the weights, and F_{x_i} are cdfs of x which are invertible.

Need N from x

Algorithm

Step 1: Generate $u \sim unif(0,1)$

Step 2: Use discrete transform to generate 1

$I - P(I = i)$

$1 - P_1$

$2 - P_2$

$n - P_n$

Step 3: Generate $v \sim unif(0,1)$

Step 4: Generate x using $F_X^{-1}(v)$

Example:
$$F_X(x) = \frac{x^2}{4} + \frac{x^3}{4} + \frac{e^x - 1}{2(e-1)}$$

$$F_{X_1}(x) = x^2 \rightarrow F_{X_1}^{-1}(v) = \sqrt{v}$$
$$F_{X_2}(x) = x^3 \rightarrow F_{X_2}^{-1}(v) = \sqrt[3]{v}$$
$$F_{X_3}(x) = x^4 \rightarrow F_{X_3}^{-1}(v) = \sqrt[4]{v}$$

Repeat the following n times:

1. Generate $u(u \leftarrow runnif(1))$??

88

2.

$$I = \begin{cases} 1 & 0 \le u \le 1/4 \\ 2 & 1/4 \le u \le 1/2 \\ 3 & 1/2 \le u \le 1 \end{cases}$$

3. Generate $v \leftarrow runif(1)$??

$$f_X(x) = \frac{2x}{4} + \frac{3x^2}{4} + \frac{e^x}{2(e-1)}$$

MATLAB pseudo code:

```
u = rand(1e5, 1);
I = InverseMethod(u);


run iteration based on I


if ( I[i] == 1 ) inverse method for F_X_1
if ( I[i] == 2 ) inverse method for F_X_2
if ( I[i] == 3 ) inverse method for F_X_3



%%%

n = 1;
while n < num+1
   if I(n) == 1
       X(n) = v(n)*0.5 % inverse of Fx1
   elseif I(n) == 2
       X(n) = v(n)*1/3 % inverse of Fx2
   elseif I(n) == 3
       X(n) = log((exp(1)-1)*v(n)+1) % inverse of Fx3
   end

   n = n + 1
```

```
hist(X, 100)
```

G   Review

Normal, uniform, exponential distributions only ones to consider. Given PDFs of all of them. Be able to generate random variables, compute things from PDFs.

Look at formula sheet.

1. Fundamentals of probability – Reubens (20% of final)

    (a) know how to define event spaces for flipping coins, rolling die, stuff

    (b) mutually exclusivity

    (c) dependence

    (d) total probability rule

    (e) conditional probability

2. Common distributions (normal, exp, unif) – work with these to calculate things like

    (a) mean

    (b) variance

    (c) covariance given two random variables Var(X), Var(aX+b), Var(X+Y), Cov(X,Y).
    If X and Y are indep, then $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. If X and Y are uncorrelated if $Cov(X,Y) = 0$.
    100% on the final is multivariate normal distribution and creating other distributions from that. **Generating random variables using the uniform distribution.**

3. Transformation of random variables. **Apply these transformations given transformation method.** Derivation unnecessary. Derivation of jacobian unnecessary. How to apply/when to apply.

(a) Univariate (!) – apply when 1d

(b) Multivariate – apply when multi d

(c) Probability integral transform – useful for constructing samples of random variables – particuarly uniform

4. Generate samples of random variables – don't know how to derive, but how to implement

(a) **Inverse transform** – useful for any discrete, for continuous if can construct inverse can construct inverse transform — 1) generate uniforms between 0-1, 2) $F_X^{-1}(U) = X$ – inverse cdf, works because of probability integral transform, if you take values from X and appyly cDF of X, then resulting thing is uniform of 0,1

(b) Polar algorithm – 0 mean, unit variance – can then construct anything from it (effeen transformation?) – useful for generating any gaussian dist

(c) **Compositional** – sums of weighted distributions, two steps 1) generate the uniform, 2) see where uniform falls , then break it down into different ps 3) get new uniform from where unifrom falls on scale, (what p it gets), then feed U back into the $F_{X_{yourp}}$. $F_X(x) = p_1 F_{x_1}(x) + ... + p_n F_{X_n}(X)$ generating number then generate

(d) Acceptance/Rejection Method – if can't use the above three, must use this. Last one you should try if none are applicable. Wastes generated samples of the uniform, less efficient than above. Most sophisticated to try to construct sampling because must figure out how to bound pdf with other pdf. – seems to imply not going to be on the exam

(e) How to apply each method or comb. of methods to generate samples. If normal, use polar with properties of efeen. Combining inverse and polar to solve problems very likely.

(f) Generating samples for Markov Chains – inverse transform where current dist. is current state – for each unique state have different inverse transform algorithm embedded

Markov chain has two things: one is inital distribution $\pi_0 = [0.30.50.2]$, matrix A is N by M, where M is number of states. Proabibilities are

the probability from what state to what state. Where tos are the rows and froms are the columns. Each row has to add to 1.

Generate state one with initial distribution. Generate uniform then look where uniform falls, if its from 0 to 0.3 it's x[1]=1, if it's from 0.3 to 0.8 it's x[1]=2, if it's from 0.8 to 1 x[1]=3. This is like markov chains.

Generate state 2. if x[1]=1, get uniform, then look at first row of matrix A. If between 0 and 0.1, then x[2]=1, ...

5. Random process – know definitions and how to identify, markov and other one,

   (a) Independent and IID – no memory, time-invariant/identically distributed
       i. If you can see these properties, just state it's an IID by inspection

   (b) **Markov Chains** (2 questions on markov chains) – conditional probability, X[k] depends on X[k-1]
       i. State space
       ii. initial state distribution
       iii. Transition probability matrix – all rows must sum to one, because each row has own probaiblity distribution.
       iv. Chapman-Kolmogorow Equation – at any time can evaluate the distribution based on initial distribution – A' is the transpose of the transition matrix – **COMMIT TO MEMORY**
          Not interested in generating anything.

$$\pi_1 = A^T \pi_0$$

$$\pi_2 = A^T \pi_1 = (A^T)^2 \pi_0$$

$$\pi_i = (A^T)\pi_0 = A^T \pi_{i-1}$$

       v. If regular markov chain, it's guaranteed that a **stationary distribution** exists: $\pi[\infty] = A'\pi[\infty]$ $'\pi[\infty] = 1$ – **warning** cannot generate random samples of markov chain using stationary distribution. If you're not using transition matrix then you're doing it wrong.

92

stationary distribution — property markovs chains have, $\pi_\infty = (A^T)^\infty \pi_0 = (A^T - I)\pi_\infty = 0$ Where I is the identitiy. Solve system of equations for π_{∞_1}, π_{∞_2}, π_{∞_3}, also add other other equation where $\pi_{\infty_1} + \pi_{\infty_2} + \pi_{\infty_3} = 1$. Then π_∞ is a vector. It basically shows how much time you spend in each state. Not all markov chains have stationary distribution – only regular. It has to be aperiodic (doesn't get stuck in infinite loops), irreducable (you can get from anywhere to anywhere). It's regular if you can find such a k, if you can take A^k and all entries are nonzero.

  vi. **Aperiod** recurrance time is null parameter – do not have regular markov chain

  vii. **Recurrance** relation – can travel in finite steps back to previous state, hard to check

(c) Weak law of large numbers – bridges between statistics and probability theory: probabilistic mean = statistical mean with many samples **mean of random process**, auto-covariance, cross-covariance

6. Bayesian inference

(a) **Bayes rule, conditional probability, marginal**

(b) **Maximum likelihood** – maximize maximum function

(c) **Bayesian maximum** – maximizes likelihood multiplied by the prior

(d) **Conditional expectation** – definition, distiribution is now conditional distribution, optimal in mean-square error sense (estimate is optimate in mean-sequare errer sense)

(e) **Law of Iterated Expectation** – useful, if take expectation of expecation of X given Y is expectation of X

(f) **Recursive Bayesian Estimation** – either remember formula or do derivation, if you want estimate of x where it doesn't depend on time, you can do summation of Probabilities. It's in the slides, expand out using bayes rule. **Will ask for a recursive basian estimatain formula.**

H   Review II

H.1   Example 1 – know multivariate and univariate

Consider multivariate normal distribution for $X = [X_1, X_2, \ldots X_n]'$

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^2 \det \Sigma}} \exp\left(-\frac{1}{2}(x - \mu)\Sigma^{-1}(x - \mu)'\right)$$

"Main task is to find B and A, then use Bayes rule and ... to find distributions" – something at this example he said "commit to memory for final".

H.2   Example 2 – uncorrelated random variables

If independent, guaranteed to be uncorrelated. If uncorrelated, not necessarily independent. Prove this just through counterexample.

Consider where $X \sim N(0, 1)$ and $Y = X^2$:

From $Y = X^2$:

$$Cov(X, Y) = E[XY] - E[X]E[Y] = \underset{0}{\cancel{E[X^3]}} - \underset{0}{\cancel{E[X]}}E[X^2] = 0$$

So uncorrelated but not independent.

Only for gaussian distributions, if uncorrelated then independent. '

H.3 Example 3 – Markov Process

I Final Review Examples

I.1 HW5 Example 2 - Resistor Selection

Suppose we have a resistor R from a batch with parameters $\mu = 1000\Omega$ and $\sigma = 200\Omega$.

What's the probability that R will be in the internal $[900, 1100]\Omega$?

—

$$
\begin{aligned}
P(900 < R \leq 1100) &= P(R \leq 1100) - P(R \leq 900) \\
&= F(1100) - F(900) \\
&= 0.69146246 - 0.30853754 \\
&= 0.3829249225
\end{aligned}
$$

I.2 HW5 Example 3 - Failure Estimate

A company produces one defective chip for every five good chips. The defective chips have a time of failure X that obey the cdf:

$$
F_{X|Y}(x|Y = 0) = (1 - e^{-x/2})u(x)
$$

The time of failure for the good chips obey the cdf:

$$
F_{X|Y}(x|Y = 1) = (1 - e^{-x/10})u(x)
$$

(a) Compute the cdf $F_X(x)$ (b) If you purchased a chip from this company, what is the probability that the chip will fail before six months of use?

—

Good chips, $Y = 0$:
$$P(Y = 0) = \frac{1}{6}$$

Bad chips, $Y = 1$:

$$P(Y = 1) = 1 - P(Y = 0) = 1 - \frac{1}{6} = \frac{5}{6}$$

(a):

$$F_X(x) = F_{X|Y}(x|Y = 0)P(Y = 0) + F_{X|Y}(x|Y = 1)P(Y = 1)$$
$$= \left[\frac{1}{6}(1 - e^{-x/2}) + \frac{5}{6}(1 - e^{-x/10})\right]u(x)$$

(b): Probability of failure before 5 months:

$$P(X \leq 6) = F_X(6) = \frac{1}{6}(1 - e^{-6/2}) + \frac{5}{6}(1 - e^{-6/10}) = 0.534$$

HW5 Example 4 - Conditional Poisson

Let X be a Poisson random variable with a rate parameter $\lambda > 0$,

$$f_X(x) = \frac{\lambda^x}{x!}e^{-\lambda} \qquad x \in \{0, 1, \dots\}$$

Let Y be the event that $\{X \text{ is even}\}$ Compute (a) the pdf $f_X(x|Y)$ and (b) the cdf $F_X(x|Y)$

—

By Baye's rule:
$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

$$f_Y(y) = P(\{X \text{ is even}\}) = \sum_{x=0,2,\dots}^{\infty} \frac{\lambda^x}{x!}e^{-\lambda}$$

And we'll define

$$f_Z(z) = P(\{\, X \text{ is odd}\,\}) = \sum_{x=1,3,\dots}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda}$$

By the property of factorial sums:

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} (-1)^x = e^{-\lambda}$$

$$f_Y(y) - f_Z(z) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} (-1)^x = e^{-2\lambda}$$

And so $f_Y(y) = \frac{1}{2}(1 + e^{-2\lambda})$

$$f_{XY}(x,y) = P(X = x, X \text{ is even})$$

I.3 Multivariate Normal Distribution

Random vectors not random variables. Let's assume you have three

No information theory on final.

Derive conditional distributions.

Define marginal distributions.

Transformations of multivariate normals.

Conditional probabilities/expectations.

Random processes.

Recursive Baysians.

Markov Chains – learn from slides and tutorial notes.

Solve markov chains using R – general algorithm/pseudocode.

Compute **stationary conditions**. How to compute **stationary distribution**. **Chapman-Kolomogorov** equation. Need to **write pseudocode to sample from a markov chain**.

Assignment 3, estimate matrix A – must know. Generate sequence of states, then count number of transitions and transitions into other states.

Use polar when sampling normal.