Probability Theory PhL 7.24.13

1. What *is* a Random Variable?

I have an appendix in Spectral on this subject, and another one in Scrambler (at least right now), and I keep mentioning the phrase "random variable" without knowing what it really means. I have two different dimensions of averaging, horizontal across a sequence or pulse train, and vertically for an ensemble of pulse trains, and these get mixed together.

I downloaded Proakis 4th Ed Digital Communications (2000). I have to add a text index to it as an evening task sometime.

There is now a 5th Ed (2007) which I don't have.

Let's now review his Chapter 2 on Probability stuff.

2.1 Probability. Rolling a die, possible outcomes, mutually exclusive events (die 2 or 4, or die 1 3 or 5), union, events

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i) \qquad (2\text{-}1\text{-}7)$$

Then joint probability idea, then conditional as I have it.

Page 21. Statistical independence. He says if P(A,B) = P(A)P(B) then these events are statistically independent. (which seems to conflict with what I have said?) He just says you can "extend" the idea of SI to 3 or more factors

$$P(A_1, A_2, A_3) = P(A_1)P(A_2)P(A_3)$$

In the general case, the events A_i, $i = 1, 2, \ldots, n$, are statistically independent provided that the probabilities of the joint events taken $2, 3, 4, \ldots,$ and n at a time factor into the product of the probabilities of the individual events.

A little hazy.

2-1-1 Random Variables Etc.

Sample space = set of values event can take I think.

Random variable is a FUNCTION mapping S to reals. X(s)

For coin toss, X(s) = 1 if heads, X(s) = -1 if tails, say.

Then extends idea to a continuous sample space.

$$F(x) = P(X \leq x) \quad (-\infty < x < \infty) \tag{2-1-19}$$

The function $F(x)$ is called the *probability distribution function* of the random variable X. It is also called the *cumulative distribution function* (cdf). Since

and then later

The derivative of the cdf $F(x)$, denoted as $p(x)$, is called the *probability density function* (pdf) of the random variable X. Thus, we have

$$p(x) = \frac{dF(x)}{dx} \quad (-\infty < x < \infty) \tag{2-1-20}$$

or, equivalently

$$F(x) = \int_{-\infty}^{x} p(u)\, du \quad (-\infty < x < \infty) \tag{2-1-21}$$

Let's stop right here and try to translate this into my little scrambler world.

Imagine a horizontal sequence of values. If it is just a set of fixed numbers, it would seem there is no probability theory involved at all.

sequence = ...1,0,3,-1,2,....

In such a sequence there is some mean value like

$<a_n> = \Sigma_n\, a_n\ N(a_n)/N\ = \Sigma_n\, a_n\ p(a_n)$

Yes, you can say that $p(a_n)$ is the probability within this sequence that a symbol has value a_n. But so far it is not a "random sequence", it is just a sequence.

You could think of this as an ensemble horizontally of single symbol events. Like each symbol is a coin toss. Then the fixed sequence above is just one possible "experiment". If you write down a list of experiments, then you have my vertical ensemble of sequences.

Go back to the single sequence as a series of separate 1-symbol events. Then you can think of the horizontal sequence as an ensemble as well, an ensemble of single sequence events. Just saying this does not make those events "independent". The events are n = 1,2,3.... and values are a_n. Then you can talk about $E(a_n)$ in the horizontal sense. Once you have some kind of ensemble, then you can apply all of my theory to it. But here we are talking ensemble of single-symbol events. Then my line above makes sense

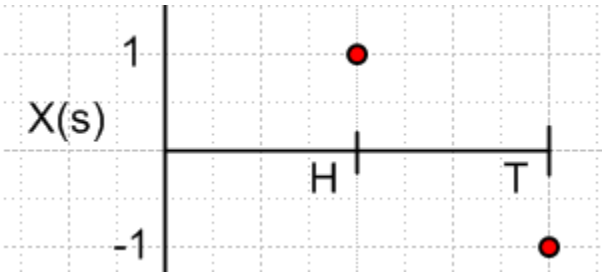
$<a_n> = \Sigma_n\, a_n\ N(a_n)/N\ = \Sigma_n\, a_n\ p(a_n)\ = E(a_n)$

Now what is the "random variable" in this situation? Well, the sample space S is perhaps GF(p) or some other discrete set of allowed values, perhaps just {1,0}.

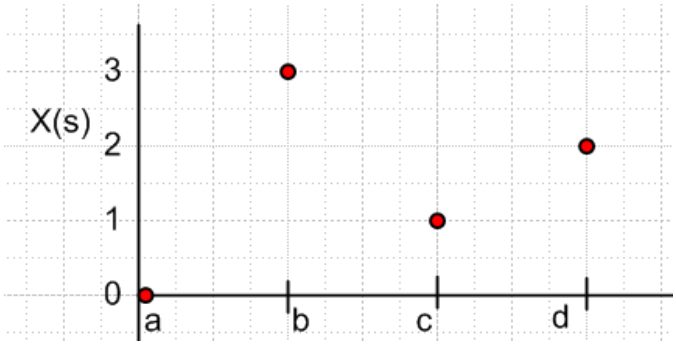
Go back to random variable X(s) for coin toss

$$X(s) = \begin{cases} 1 & (s = H) \\ -1 & (s = T) \end{cases} \qquad (2\text{-}1\text{-}18)$$

Here is my plot of this function whose domain is discrete.

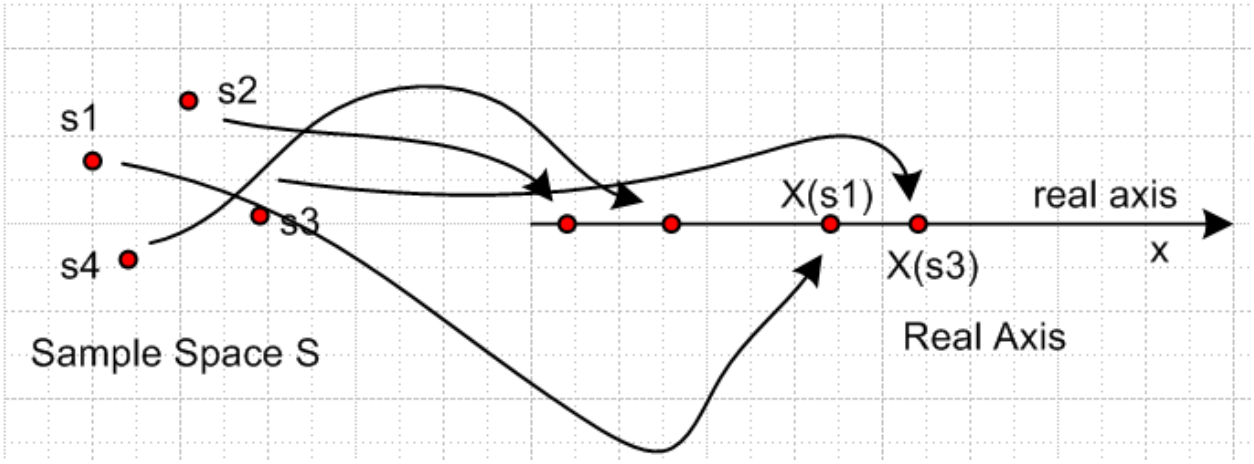


I do NOT understand what this means. It seems to mean nothing. There is no "probability" involved in this random variable definition. For GF(2^2) you could write this as



where maybe the horizontal axis represents *abstract elements* of GF(2^2) while the vertical axis is *real numbers* in this case integers.

Basically it is a way to get from whatever your "events are " to a set of numbers. Here is another picture



But in MY case, we have $s_1 = 0$, $s_2 = 1$ etc, and the mapping is then from reals to reals and it is just an identity mapping.

X(s) = s or change variable name X(x) = x

So my "random variable" is the function X(x) = x. Very boring. Maybe write this $X(a_n) = a_n$. So this seems strange. How can you define a "random variable" to be a function? It does not make sense .

Go to another source. Wiki has a completely different definition

> In probability and statistics, a **random variable** or **stochastic variable** is a variable whose value is subject to variations due to chance (i.e. randomness, in a mathematical sense). As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown); rather, it can take on a set of possible different values, each with an associated probability.

So at least they define it as a variable! But later they say

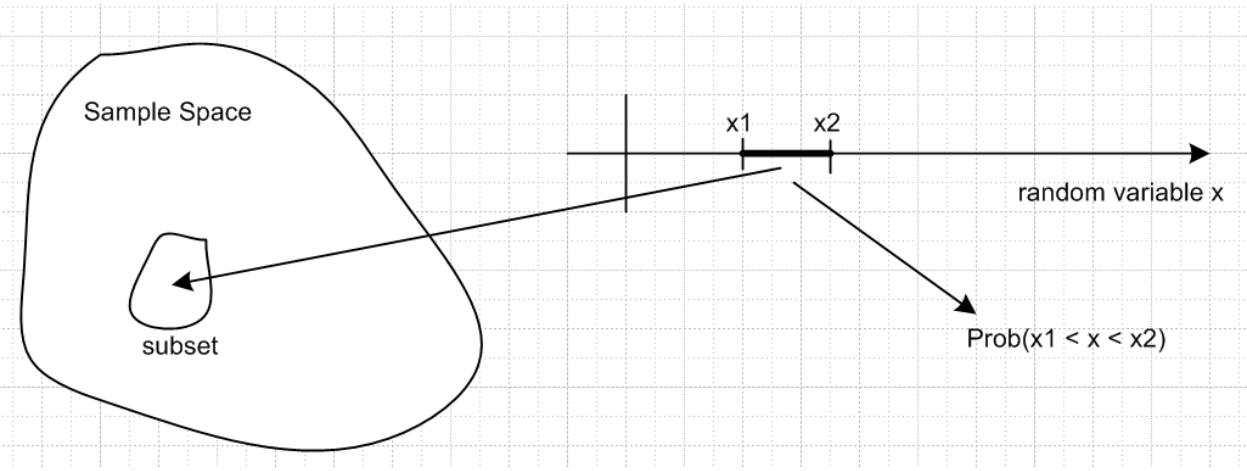
> *Random variable* is usually understood to mean a real-valued random variable; this discussion assumes real values. A random variable is a real-valued function defined on a set of possible outcomes, the sample space Ω. That is, the random variable is a function that maps from its domain, the sample space Ω, to its range, the real numbers or a subset of the real numbers. It is typically some kind of a property or measurement on the random outcome (for example, if the random outcome is a randomly chosen person, the random variable might be the person's height, or number of children).

So here the set S is a set of "outcomes", still the sample space. In their example, the sample space might be a set of people including Phil. So Phil is a possible "outcome" in this space. The "variable" might be the height of Phil which is a real number, though Phil is not a real number. So here the random variable is the height of a person. The randomness enters because you have select a person from your ensemble of

people. Then you might ask what is the probability that for this ensemble height is in range h to h+dh and that would be a pdf.

Maybe I like to thing of the "outcomes" as the "possible events". Then Phil is a possible event in the group of people. The events need not be numbers.

They go on to make a good comment:



For any range of the "random variable" x (height of person), you have to be able to associate any region of the real line (x_1,x_2) with a well-defined subset in the sample space, that is, with a particular subset of the 10,000 people in the Sample Space.

But then they refer to the subsets as "events". So an event could be a group of people.

The random variable X(s) is like X(Phil) = 6'1".

A PDF for a discrete value is called a Probability Mass Function or PMF of the values are non-negative integers like the number of children of a person. Then X(phil) = 0. That type of random variable is called a PMF. Fine.

Wiki goes on to say:

> In examples such as these, the sample space (the set of all possible persons) is often suppressed, since it is mathematically hard to describe, and the possible values of the random variables are then treated as a sample space. But when two random variables are measured on the same sample space of outcomes, such as the height and number of children being computed on the same random persons, it is easier to track their relationship if it is acknowledged that both height and number of children come from the same random person, for example so that questions of whether such random variables are correlated or not can be posed.

So values of height h become the new sample space and the space of people is "suppressed".

Let's try now to interpret the coin toss discussion.

The possible outcomes for one coin toss can be described by the sample space Ω = {heads, tails}. We can introduce a real-valued random variable Y that models a $1 payoff for a successful bet on heads as follows:

$$Y(\omega) = \begin{cases} 1, & \text{if } \omega = \text{heads}, \\ 0, & \text{if } \omega = \text{tails}. \end{cases}$$

If the coin is equally likely to land on either side then Y has a probability mass function f_Y given by:

$$f_Y(y) = \begin{cases} \frac{1}{2}, & \text{if } y = 1, \\ \frac{1}{2}, & \text{if } y = 0, \end{cases}$$

Since the random variable maps the two events heads and tails into a set of non-negative integers, the probability function gets that special name PMF. Fine. Now for the first time we have some kind of "probability function" associated with heads and tails. It need not be 1/2 and 1/2 as shown. I would probably just call this thing a pdf anyway, a discrete PDF.

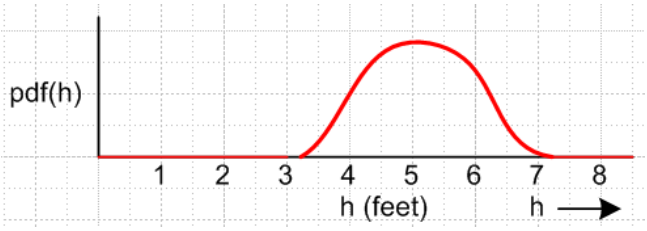
So again we have this indirection of outcomes → real numbers → probability. Y(w) is the random variable here.

The meaning of a Random Variable

Before attempting a definition, we give a few examples.

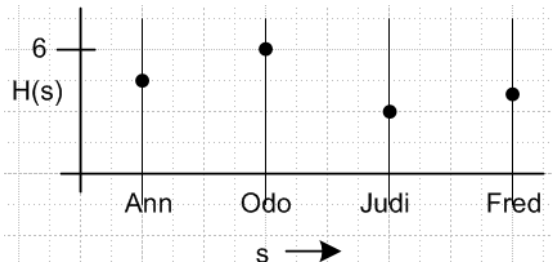
One example of a **random variable** is the height h of a person, where the person is one of many people in some ensemble (set) of people, each of whom has some well-defined height. This ensemble of people is an example of a **sample space** and the individual people are the **samples**. Any subset of people forms an **event**, so {Joe} is an event, and {Joe, Alice} is a different event. This h is a "real variable" because it can take various real values. In this sample space the people are not real numbers, they are people, but h is a real number. In this example, the random variable h is continuous whereas the sample space is discrete.

If we measure the height of all the people in the ensemble, we end up with some probability density function (pdf) for height, where the area under the red curve is 1 :



This pdf would be different for a set of NBA basketball players versus a set of horse racing jockeys. The pfd meaning is that the probability of someone in the sample space having a height between h and h+dh is given by pdf(h)dh.

Clearly height in this situation can be regarded as a mapping from the sample space S to the real numbers R. We might call this mapping H, so H:S→R. This mapping is a function whose domain is S and whose range lies in R. We write then h = H(s), with an example 6'1" = H(Joe). For a discrete sample space one could place the samples of the sample space in some sequential order and make a picture of H(s) like so



It is from the function H(s) that one computes pdf(h) by tossing people into narrow height bins, and then plotting the bin counts divided by the total number of people.

Another example of a random variable would be temperature at all points on a 2D surface. The sample space here is the continuous 2D surface, and a sample is a point (x,y) on that surface. Temperature is a real number, but the samples are not real numbers, they are 2D vectors. The mapping is t = T(s) where s = (x,y), so t = T(x.y) and we could plot this as a 2D surface lying over the x,y planar sample space.

Now we attempt a definition.

Definition: **Random Variable**. A random variable x over S is some real quantity x which has a probability distribution with respect to some particular sample space S. We have seen above where the word "variable" applies, since our "real quantity x" is a variable which takes some set of real values. The word "random" is associated with the fact that the variable x doesn't have some fixed value, but has a value that is described by a pdf. There is some probability distribution pdf(x) associated with variable x which can be computed from x = X(s). That distribution might have a Gaussian shape and if it did, one might call x a "Gaussian random variable". But the raw definition does not specify any particular probability distribution, it just requires that there be one. Nor does the definition specify a particular sample space.

In both the height and temperature examples the random variables h and t were continuous so the pdf(x) was a continuous function.

One can of course have discrete random variables and the classic example involves tossing coins. One can consider a sample space S not of people, but of coins which have just been tossed. Just as a person has a property called height, a tossed coin has a property called heads or tails. A difference is that this property is not a real number, so in order to talk about a random variable, we have to associate perhaps heads with 1.7 and tails with 2.6 and call this real variable h. As before, h = H(s) and we can write down the value of h for each tossed coin in the sample space, and from this we can construct a pdf which might look like this:



Here either variable h or function H(s) would be called a random variable. This h is a real quantity which has a probability distribution relative to some sample space S (a set of tossed coins). This is a discrete random variable. A discrete pdf function is often called a pmd meaning a **probability mass distribution**, suggested by the idea that the discrete random variable values are like point particles with all their mass concentrated at a point, rather than like distributed matter having a continuous mass density.

If the coins were equally weighted on both sides, and if S contained 10^6 such tossed coins, we would expect the heights of the two red lines to be the same and to be 1/2. Maybe the coins are asymmetric so as to produce the 2/3, 1/3 distribution shown, or perhaps the tosses were not "honest". One might imagine an honest toss of an equally balanced coin, and one might call this a "random coin toss". But even it tosses are not random in this equal probability sense, h = H(s) is still called a "random variable" because, no matter how the coins were tossed, there does exist some H(s) and some pdf(h). This seems to be a point of confusion regarding random variables. Random just means "there is some associated probability distribution".

> The formal mathematical treatment of random variables is dealt with in the subject of probability theory. In that context, random variables are defined in terms of functions defined on a probability space. By this interpretation, "When we speak of a random variable, we think in addition of the probability distribution according to which it takes all possible values. As a result, a random variable is misnamed as it is neither random nor a variable."[1]

****

How does a rolling die fit into this framework? The sample space could be a set of N dice that have just been rolled. A die is not a real number. The random variable could be the number facing up. That variable, like height, is already a real number. We have some random variable n = N(s) which enumerates how the dice have all come out. There is some associated pdf(n) we could construct. It might be a set of 6 bars all of height 1/6 or some other set of bar heights for an uneven die.

Now where would the sum of two die values fit in? Ouch, not working very well. My model is not really very good. I think the numbers {1,2,3...6} should be the sample space, not the die. And for coin toss, the sample space should be {heads,tails}. And what about the word outcome? Is the person's height an outcome? The wiki site is very fuzzy on this. I would like to fit many different examples into the a model that works for all.

Let's now look at a different source.

Definition 2.6.1 *Let* **E** *be an experiment whose outcomes are a sample space* $\mathcal{U}$ *for which the probability* $P(e)$ *is defined for each outcome* $e \in \mathcal{U}$. *A random variable is a function* $X(e)$ *that associates a number with each outcome* $.e \in \mathcal{U}$.

Here "outcomes" are elements of the "sample space". But still very vague. Author defines an "event" as being a subset of S which makes to an interval on the reals. Each interval maps back into some subset we know that, but not every subset maps into an interval. The ones that map into intervals are "events". This seems to me to be a rather strange definition of an "event". Why is that word chosen? We need examples. These authors all went to get on to things like CDF and PDF, they don't want to define their objects. A common problem.

trated as follows: to each experiment that we consider there will be associated a random variable, which represents the outcome of any particular experiment. The set of possible outcomes is called the *sample space*. In the first part of this section,

This does not fit well with the idea of sample space = set of people. If we have one die, then the sample spacing being {1.2.3.4.5.6} does fit with the above. It is the set of outcomes of an experiment. There is only one die. Continuing

We now make two additional definitions. These are subsidiary to the definition of sample space and serve to make precise some of the common terminology used in conjunction with sample spaces. First of all, we define the elements of a sample space to be *outcomes*. Second, each subset of a sample space is defined to be an *event*. Normally, we shall denote outcomes by lower case letters and events by capital letters.

Example 1.6 A die is rolled once. We let X denote the outcome of this experiment. Then the sample space for this experiment is the 6-element set

$$\Omega = \{1, 2, 3, 4, 5, 6\} ,$$

where each outcome i, for $i = 1, \ldots, 6$, corresponds to the number of dots on the face which turns up. The event

$$E = \{2, 4, 6\}$$

corresponds to the statement that the result of the roll is an even number. The event E can also be described by saying that X is even. Unless there is reason to believe the die is loaded, the natural assumption is that every outcome is equally likely. Adopting this convention means that we assign a probability of $1/6$ to each of the six outcomes, i.e., $m(i) = 1/6$, for $1 \leq i \leq 6$. □

The "experiment" involves one die. The sample space Ω is as shown. An event is shown. The random variable here might be n = N(ω) = ω, pretty simple. There is some pdf(ω). "A random variable represents the outcome of an experiment. " Strictly N(ω) is the random variable here.

Example 1.7 Consider an experiment in which a coin is tossed twice. Let X be the random variable which corresponds to this experiment. We note that there are several ways to record the outcomes of this experiment. We could, for example, record the two tosses, in the order in which they occurred. In this case, we have $\Omega = \{\text{HH,HT,TH,TT}\}$. We could also record the outcomes by simply noting the number of heads that appeared. In this case, we have $\Omega = \{0,1,2\}$. Finally, we could record the two outcomes, without regard to the order in which they occurred. In this case, we have $\Omega = \{\text{HH,HT,TT}\}$.

The "experiment" is two coin tosses. Random variable X represents the outcome of this experiment. The same space is not Ω = {H,T} it is sort of {H,T}⊗{H,T} which author writes in several ways. So here we get away from the idea that Ω = {H,T} for any number of coins being tossed.

An "event" here might be getting 2 heads. This is the same as HH in Ω. But the "event" of getting only one head is HT + TH. An event might be a single die rolling an even number.

Example 1.9 Three people, A, B, and C, are running for the same office, and we assume that one and only one of them wins. The sample space may be taken as the 3-element set $\Omega = \{\text{A,B,C}\}$ where each element corresponds to the outcome of that candidate's winning. Suppose that A and B have the same chance of winning, but that C has only 1/2 the chance of A or B. Then we assign

$$m(\text{A}) = m(\text{B}) = 2m(\text{C}) \ .$$

Since

$$m(\text{A}) + m(\text{B}) + m(\text{C}) = 1 \ ,$$

The "experiment" I guess is an election. The three outcomes are "A wins", "B wins", or "C wins", and that is what is meant by Ω = {A,B,C}.

4. If A and B are *disjoint* subsets of Ω, then $P(A \cup B) = P(A) + P(B)$.

Theorem 1.2 If $A_1, \ldots, A_n$ are pairwise disjoint subsets of Ω (i.e., no two of the A_i's have an element in common), then

$$P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) \ .$$

This all seems reasonable.

So how do I deal with the people and the heights? What is the experiment? It is like an experiment that has already been done and we are just seeing the outcome? Each person is like a die. So the experiment is a room of 2 people, the outcomes are what? This fits very poorly. If people could only have 6 heights, then the same space would be {123456}⊗ {123456} like the two coin tosses.

Definition 1.3 The *uniform distribution* on a sample space Ω containing n elements is the function m defined by

$$m(\omega) = \frac{1}{n} \ ,$$

for every $\omega \in \Omega$. □

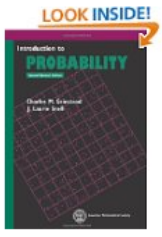
Example 1.13 A coin is tossed until the first time that a head turns up. Let the outcome of the experiment, ω, be the first time that a head turns up. Then the possible outcomes of our experiment are

$$\Omega = \{1, 2, 3, \ldots\} \ .$$

Note that even though the coin could come up tails every time we have not allowed for this possibility. We will explain why in a moment. The probability that heads

The experiment is what is explained in the first sentence.

What is this book I am looking at? Grinstead is math prof at Swarthmore College. Snell is was at Dartmouth math, he died in 2011.



Introduction to Probability: Second Revised Edition by Charles M. Grinstead and J. Laurie Snell (Oct 9, 2012)

Formats	Price	New	Used
Paperback	$60.00 **$53.70** Prime	$53.70	$49.70
Usually ships in 1 to 3 weeks			
Eligible for FREE Super Saver Shipping.			

Product Details

Paperback: 510 pages

Publisher: American Mathematical Society (October 9, 2012)

Language: English

ISBN-10: 0821894145

ISBN-13: 978-0821894149

Product Dimensions: 9.9 x 7 x 1.1 inches

1 Discrete Probability Distributions **1**
1.1 Simulation of Discrete Probabilities . . . 1
1.2 Discrete Probability Distributions . . . 18

2 Continuous Probability Densities **41**
2.1 Simulation of Continuous Probabilities . . . 41
2.2 Continuous Density Functions . . . 55

3 Combinatorics **75**
3.1 Permutations . . . 75
3.2 Combinations . . . 92
3.3 Card Shuffling . . . 120

4 Conditional Probability **133**
4.1 Discrete Conditional Probability . . . 133
4.2 Continuous Conditional Probability . . . 162
4.3 Paradoxes . . . 175

5 Distributions and Densities **183**
5.1 Important Distributions . . . 183
5.2 Important Densities . . . 205

6 Expected Value and Variance **225**
6.1 Expected Value . . . 225
6.2 Variance of Discrete Random Variables . . . 257
6.3 Continuous Random Variables . . . 268

7 Sums of Random Variables **285**
7.1 Sums of Discrete Random Variables . . . 285
7.2 Sums of Continuous Random Variables . . . 291

8 Law of Large Numbers **305**
8.1 Discrete Random Variables . . . 305
8.2 Continuous Random Variables . . . 316

9 Central Limit Theorem **325**
9.1 Bernoulli Trials . . . 325
9.2 Discrete Independent Trials . . . 340
9.3 Continuous Independent Trials . . . 355

10 Generating Functions **365**
10.1 Discrete Distributions . . . 365
10.2 Branching Processes . . . 377
10.3 Continuous Densities . . . 394

11 Markov Chains **405**
11.1 Introduction . . . 405
11.2 Absorbing Markov Chains . . . 415
11.3 Ergodic Markov Chains . . . 433
11.4 Fundamental Limit Theorem . . . 447
11.5 Mean First Passage Time . . . 452

12 Random Walks **471**
12.1 Random Walks in Euclidean Space . . . 471
12.2 Gambler's Ruin . . . 486
12.3 Arc Sine Laws . . . 493

Appendices **499**
A Normal Distribution Table . . . 499
B Galton's Data . . . 500
C Life Table . . . 501

Index **503**

I inadvertently downloaded this entire book! So I now save it as perhaps my only book on this subject. It has a nice history section at the end of Chapter 1.

It is frequently useful to be able to refer to an outcome of an experiment. For example, we might want to write the mathematical expression which gives the sum of four rolls of a die. To do this, we could let X_i, $i = 1, 2, 3, 4$, represent the values of the outcomes of the four rolls, and then we could write the expression

$$X_1 + X_2 + X_3 + X_4$$

Here experiment = roll four dice, outcome Ω = possible sums of all four = {4,5.....36}, and this would be the sample space.

I have done a reasonable scan of the first chapter on "discrete probability distributions", and now will look into Chapter 2 on continuous ones. He opens with the "spinner" but I don't get it! I think the possible outcomes here are an angle 0 to 360 which he regards as (0,1). Fine. The outcome is continuous.

Darts

Example 2.8 A game of darts involves throwing a dart at a circular target of *unit radius*. Suppose we throw a dart once so that it hits the target, and we observe where it lands.

To describe the possible outcomes of this experiment, it is natural to take as our sample space the set Ω of all the points in the target. It is convenient to describe these points by their rectangular coordinates, relative to a coordinate system with

More generally, we suppose that suitable coordinates can be introduced into the sample space Ω, so that we can regard Ω as a subset of $\mathbf{R}^n$. We call such a sample space a *continuous sample space*. We let X be a random variable which represents the outcome of the experiment. Such a random variable is called a *continuous random variable*. We then define a density function for X as follows.

He then does the usual pdf stuff. But he never talks about an experiment with two spinners. I guess the sample space would be a subset of R^n as he says above. So for a room with 10 people, the Ω would be a chunk of R^{10} with each real axis between 5 and 6 feet?

He gets on to the idea of "a function of a random variable".

Suppose that X is a discrete random variable with sample space Ω, and $\phi(x)$ is a real-valued function with domain Ω. Then $\phi(X)$ is a real-valued random variable. One way to determine the expected value of $\phi(X)$ is to first determine the distribution function of this random variable, and then use the definition of expectation. However, there is a better way to compute the expected value of $\phi(X)$, as demonstrated in the next example

Theorem 6.1 If X is a discrete random variable with sample space Ω and distribution function $m(x)$, and if $\phi : \Omega \to \mathrm{R}$ is a function, then

$$E(\phi(X)) = \sum_{x \in \Omega} \phi(x) m(x) \ ,$$

provided the series converges absolutely. □

The proof of this theorem is straightforward, involving nothing more than grouping values of X with a common Y-value, as in Example 6.6.

Another source now: http://www.cut-the-knot.org/Probability/SampleSpaces.shtml

A *sample space* is a collection of all possible outcomes of a random experiment. A *random variable* is a function defined on a sample space. We shall consider several examples shortly. Later on we shall introduce probability functions on the sample spaces. A sample space may be *finite* or *infinite*. Infinite sample spaces may be *discrete* or *continuous*.

Human height. The experiment is to randomly select a human and measure his or her length. Depending of how far reaching our means of selection is it is possible to consider a sample space of about 6.6 billion humans inhabiting the planet Earth. In this case, the height of the selected person becomes a random variable. However, it is also possible to consider the sample space consisting of all possible values of height measurements of the world population. The tallest man ever measured lived in the United States and had a height of 272 cm (8'11"). The height of the shortest person is more difficult to determine. Zero is clearly the low bound, but, for a living adult, it may be safely raised to, say, 40 cm. This suggests a sample space which is a line segment [40, 272] in centimeters. While at all times the human population is discrete, we may assume that in some height range near the normal average, all possible heights are realized making a continuous classification. Still, very certainly at the top end there are gaps and unique measurements making the upper part of the range rather discrete.

(More examples of continuous sample space can be found elsewhere.)

Very confused! First, the "sample space" is a set of humans and height is a random variable. But then author says "sample space" is line segment [40,272] cm. So which is it? The second is more in line with a set of outcomes of the experiment of measuring height.



Sample Space

The sample space is an exhaustive list of all the possible outcomes of an experiment. Each possible result of such a study is represented by one and only one point in the sample space, which is usually denoted by S.

Examples
Experiment Rolling a die once:
Sample space S = {1,2,3,4,5,6}
Experiment Tossing a coin:
Sample space S = {Heads,Tails}
Experiment Measuring the height (cms) of a girl on her first day at school:
Sample space S = the set of all possible real numbers

This reinforces the idea of samples space = outcome = height.

Sample space

In probability and statistics the term experiment is used in a very wide sense and refers to any procedure that yields a collection of outcomes. A **random experiment** is one whose outcomes depend only on chance and it can be repeated under identical conditions.

The set whose elements are all the possible outcomes of an experiment is called the **sample space** of the experiment. The concept of the sample space plays a fundamental role in determining probabilities for individual outcomes of a random process. Your understanding of the sample space should be that of **a complete and as detailed as possible listing of all the basic (elementary) outcomes of a random experiment**.

1. Random variable

   The outcome of a process that cannot be predicted with certainty.

2. Sample space

   The set of all outcomes of a given random variable.

Note that the "true" outcome of most experiments is not a number, but a physical situation, e.g., "3 dots up" or "the subject chose the blue toy". For convenience sake, we often "map" the physical outcomes of an experiment to integers or real numbers, e.g., instead of referring to the outcomes 1du to 6du, we can refer to the numbers 1 to 6. Technically, this mapping is called a **random variable**, but more commonly and informally we refer to the unknown numeric outcome itself (before the experiment is run) as a "random variable". Random variables commonly are represented as upper case English letters towards the end of the alphabet, such as Z, Y or X. Sometimes the lower case equivalents are used to represent the actual outcomes after the experiment is run.

A few more basic definitions are worth learning at this point. A random variable that takes on only the numbers 0 and 1 is commonly referred to as an **indicator (random) variable**. It is usually named to match the set that corresponds to the

2 Random Variables

Probability is the study of random variables. A random variable is a "probabilistic" outcome, such as a coin flip or the height of a person chosen from a population. Here, the observed value of heads or tails, or the observed height, depend on some known or unkown probability distribution. It may also sometimes be useful to model probabilistically quantities that are not "probabilistic" in this sense, for example the temperature on some future date, the temperature on some past date, or the number of times the word "streetlight" appears in a document.

Random values take on values in a *sample space*. This space may be discrete or continuous, and the space may be defined differently for different scenarios. For example, the sample space for a coin flip is $\{H, T\}$; the sample space for height might be defined as the positive real values in $(0, \infty)$; for temperature, it might be defined as real values in $(-\infty, \infty)$; for the number of occurrences of a word in a document, it might be the positive integers $\{1, 2, \ldots\}$.

There is not necessarily one uniquely "correct" sample space for a particular concept. For example, one may argue that using a sample space of $(0, \infty)$ for a person's height allows for impossibly tall people, and one might define an alternative sample space that puts a finite upper limit of the height one would measure. However, it typically doesn't matter if the sample space is "too big," for example infinite height or infinitely negative temperature (as long as the distribution on the sample space places no or very low probability on the impossible events).

Notice here the analogy between

experiment = flipping a coin outcome = heads, tails

experiment = picking a person from an ensemble outcome = height

Ensemble is called a "population"

So I think my first impression of people = sample space was just plain wrong.

If $f(X)$ is a function of a random variable X, then $f(X)$ is itself also a random variable. In the case that X is discrete, the *expectation* of $f(X)$ is a weighted average of f, where the weighting is determined by $p(x)$:

$$\mathrm{E}[f(X)] = \sum_x p(x) f(x)$$

OK, after this I think I am ready to write things up, manana.

Question: Compare again the roll of a die experiment to the height experiment. Both have outcomes that are already real numbers, so we don't need "functions" to map outcomes to numbers. If the die is a "fair die", then P = 1/6 for each number and that is the pdf or pmf as it is called. But for the height deal, if we make a "fair selection" from our group of 100 people, we don't get a uniform distribution, we get a certain

pdf based on the group. In the die case, the random act is "rolling the die". In the height case, the random act is "selecting a member of the group". In this selection case, we always think of the probability of selecting any of the 100 people is the same, so the resulting pdf is what we expect it to be. Also, the people already exist, whereas the die is just now being rolled.

In the die case there is some physical way to compute the outcome of any roll (micro physics) and it might not be a fair roll so you might not get a uniform distribution. But the heights of the group members are pre-established. You might model why each member has the height they have. Maybe each roll of the die generates a new person in the group with that outcome. Maybe each person is an "experiment" like the rolled die. It just that room full of people has known results.

Experiment:	N rolls of a die 1 roll of N dice	child counts of group of N people
Outcome:	N face-up numbers	N child counts
Sample Space:	N outcome sets {1,2,3,4,5,6} $\{1,2,3,4,5,6\}^{\mathbf{N}}$	N counts each in {0,1,2.....100}
Random Variable	n = face-up number	n = child count
Function:	N(s) = s, identity	N(s) = s, identity
PDF	uniform 1/6 if fair rolls	child count distribution

prob(θ) 0 2π

****

I finally found someone who verifies my theorem about expectation values:

5.6 Special Theorems

- Expected value of a function of random variables, $E[g(Y_1, Y_2, \ldots, Y_k)]$, equals

$$= \begin{cases} \sum_{\text{all } y_1} \sum_{\text{all } y_2} \cdots \sum_{\text{all } y_k} g(y_1, y_2, \ldots, y_k) p(y_1, y_2, \ldots, y_k) & \text{if discrete,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_k) f(y_1, y_2, \ldots, y_k)\, dy_1 dy_2 \cdots dy_k & \text{if continuous.} \end{cases}$$