My probability and statistics exercises

Evgeny Markin

2023

Contents

Preface

0.1 Notation

Some notable deviations from book's notation are presented here

1. Sometimes instead of p.d.f. and the likes of it, we write PDF
2. Countable set is defined as a set that has an injection into naturals (i.e. countably infinite and finite)

Chapter 1

Introduction to Probability

1.1 The History of Probability

1.2 Interpretations of Probability

1.3 Experiments and Events

1.4 Set Theory

Although the section is not pretty complex, some definitions can use a tad bit of rigor.

Sample space is a set of outcomes.

Event is a subspace of a sample space.

If S is a sample space, and $f : \mathcal{P}(S) \to R$ is a function such that

1. $\mathrm{range}(f) \subseteq [0, \infty)$
2. $f(S) = 1$
3. if A is a countable set of events (i.e. $K \subseteq \mathcal{P}(S)$ and $|K| \leq_c \omega$), and it is indexed by ω (i.e. $A = \{A_1, A_2, ...\}$) then $f(\bigcup_{i \in \omega} A_i) = \sum f(A_i)$

then f is called a probability measure, or just measure. Canonical representation of probability measure is Pr and above mentioned items are usually refered to as axioms of probability.

Definitions for this note were compiled from the book itself, book on measure theory (MIRA by Axler) and general notes from the internet.

Exercises in this section (or exercises similar to them) are handled in the set theory course

1.5 The Definition of Probability

1	2/5
2	0.7
3a	1/2
3b	1/6
3c	3/8
4	0.6
5	0.4
6	0.5
8	30
11a	1 - π/4
11b	0.75
11c	2/3
11d	0
14a	0.38, 0.16
14b	0.04

A little notation, related to 6:

$$Pr(A) = 0.5$$

$$Pr(B) = 0.2$$

$$Pr(A \cap B) = 0.1$$

$$Pr(A \cup B) = 0.6$$

$$Pr((A \cup B) \cap (A \cap B)^c) = P(A \cup B) - P((A \cup B) \cap (A \cap B)) = P(A \cup B) - P(A \cap B) = 0.5$$

1.5.7

If $Pr(A) = 0.4$ and $Pr(B) = 0.7$, then we follow that the maxium $Pr(A \cap B)$ is attained if $A \subset B$, in which case $Pr(A \cap B) = Pr(A) = 0.4$. The minimum is obtained if $A \cup B = S$, in which case $Pr(A \cap B) = 0.1$

1.5.9

The event that exaclty one of the events occurs can be expressed as

$$(A \cap B^c) \cup (A^c \cap B)$$

which comes from either the definition of xor, common sense or something else, depending on your preferences. Thus we follow that

$$Pr((A \cap B^c) \cup (A^c \cap B)) = Pr(A \cap B^c) + Pr(A^c \cap B) - Pr((A \cap B^c) \cap (A^c \cap B)) =$$

$$= Pr(A \cap B^c) + Pr(A^c \cap B) - Pr((A \cap A^c) \cap (B^c \cap B)) =$$

$$= Pr(A \cap B^c) + Pr(A^c \cap B) = Pr(A) - Pr(A \cap B) + Pr(B) - Pr(B \cap A) =$$

$$= Pr(A) - Pr(A \cap B) + Pr(B) - Pr(A \cap B) = Pr(A) + Pr(B) - 2Pr(A \cap B)$$

as desired (rules used in this derivitation: association of unions, $A \cap A^c = \emptyset$ and other trivial stuff)

1.5.10

$$Pr(A \cap B^c) = Pr(A) - Pr(A \cap B)$$

$$Pr(A \cap B^c) + Pr(A \cap B) = Pr(A)$$

as desired.

1.5.12

Suppose that $n > m \in N$. Then we follow that by definition

$$B_m \subseteq A_m$$

and

$$B_n \subseteq A_m^c$$

thus we follow that

$$B_m \cap B_n \subseteq A_m \cap A_m^c = \emptyset$$

thus

$$B_m \cap B_n = \emptyset$$

therefore we conclude that $B_1, B_2 ...$ are disjoint sets. Thus we follow that

$$Pr(\bigcup_{i=1}^{n} B_i) = \sum_{i=1}^{n} Pr(B_i)$$

For $n = 2$ we've got that

$$B_1 \cup B_2 = A_1 \cup (A_1^c \cap A_2) = (A_1 \cup A_1^c) \cap (A_1 \cup A_2) = A_1 \cup A_2$$

and by induction we can follow that

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i$$

thus

$$Pr(\bigcup_{i=1}^{n} B_i) = \sum_{i=1}^{n} Pr(B_i)$$

implies that

$$Pr(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Pr(B_i)$$

for $n \in N$. Given that n is arbirtary, we can follow that

$$Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Pr(B_i)$$

as desired.

1.5.13

First equation follow from induction on the result that

$$Pr(A \cup B) \leq Pr(A) + Pr(B)$$

the second equation follows from the first equation, DeMorgan laws and induction on the form

$$Pr(A \cap B) = Pr((A^c \cup B^c)^c) = 1 - Pr(A^c \cup B^c) \geq 1 - (Pr(A^c) + Pr(B^c))$$

1.5.14

$$Pr(A) = 0.34$$
$$Pr(B) = 0.12$$
$$Pr(O) = 0.5$$
$$Pr(AB) = 1 - 0.34 - 0.12 - 0.5 = 0.04$$
$$Pr(a - A) = 0.34 + 0.04 = 0.38$$
$$Pr(a - B) = 0.12 + 0.04 = 0.16$$

1.6 Finite Sample Spaces

1	1/2
2	1/2
3	2/3
4	1/7
5	4/7
6	1/4
8b	1/4

1.6.7

The possinble genotypes are Aa and aa with probabilities 1/2 and 1/2 respectively

1.6.8a

The sample space of the experiment is $\{heads, tails\} \times \{1, 2, 3, 4, 5, 6\}$,

1.7 Counting Methods

1	14
2	9000
3	120
4	24
5	5/18
6	5/324
7	0.014731
8	360 / 2401
9	1 / 20
10a	r/100
10b	r/100
10c	r/100

1.7.11

$$s(n) = \frac{1}{2}\log(2\pi) + (n + \frac{1}{2})\log n - n \approx \log n!$$

$$\log n! - \log{(n-m)!} = \log\frac{n!}{(n-m)!}$$

$$s(n) - s(n-m) = \frac{1}{2}\log(2\pi) + (n+\frac{1}{2})\log n - n - (\frac{1}{2}\log(2\pi) + ((n-m)+\frac{1}{2})\log n - m - (n-m)) =$$

$$= (n+\frac{1}{2})\log n - n - ((n-m)+\frac{1}{2})\log{(n-m)} + (n-m) =$$

$$= (n+\frac{1}{2})\log n - ((n-m)+\frac{1}{2})\log{(n-m)} - m \approx \log\frac{n!}{(n-m)!}$$

$P(n,m) = \frac{n!}{(n-m)!} = \exp(s(n) - s(n-m))$

1.8 Combinatorial Methods

1	184756
2	latter
3	equal
4	1 / 10626
5	-
6	2/n
7	(n - k - 1)/C(n, k)
8	(n - k)/C(n, k)
9	(n + 1)/C(2n, n)
10	$15/92 \approx 0.16304$
11	$1/75 \approx 0.01333$
12	$69/119 \approx 0.57983$
13	$173/1518 \approx 0.114$
14	-
15	-
16a	$48/175 \approx 0.27429$
16b	$2^{50}/C(100, 50) \approx 0$
17	$4C(13, 4)/C(52, 4) = 44/4165 \approx 0.0105$
18	$C(20, 2)^5/C(100, 10) \approx 0.0143$
19	-
20	-
21	C(365 + 7 - 1, 7)
22	-

1.8.5

Prove that

$$\frac{\prod_{4155 \leq i \leq 4251} i}{\prod_{2 \leq i \leq 97} i}$$

is an integer

$$\frac{\prod_{4155 \leq i \leq 4251} i}{\prod_{2 \leq i \leq 97} i} = \frac{\prod_{4155 \leq i \leq 4251} i}{\prod_{1 \leq i \leq 97} i} =$$

$$= \frac{\prod_{4155 \leq i \leq 4251} i}{97!} = \frac{4251!}{4154!97!} = \frac{4251!}{4154!(4251 - 4174)!} = C(4251, 4154)$$

and binomial coefficients are integers (pretty sure that we can follow that by induction in some more advanced course).

1.8.10

There are total of $C(24,10)$ possible subsets of length 10 in the space of 24. We follow that there are $C(22,8)$ ways to pick 8 normal bulbs, which is what required to pick 2 defective bulbs. Therefore the probability is

$$\frac{C(22,8)}{C(24,10)} = 15/92 \approx 0.16304...$$

1.8.12

Using the same logic as in 1.8.10, there is a possibility $\frac{C(33,8)}{C(35,10)}$ that same two guys will be in the first team, and probability of $\frac{C(33,23)}{C(35,10)}$ that they'll be in the other team. Thus the total probability is the sum of two.

1.8.14

Prove that for all positive integers n, k such that $n \geq k$

$$C(n,k) + C(n,k-1) = C(n+1,k)$$

$$\begin{aligned}
C(n,k) + C(n,k-1) &= \frac{n!}{(n-k)!k!} + \frac{n!}{(n-k+1)!(k-1)!} = \\
&= \frac{n!}{k(n-k)!(k-1)!} + \frac{n!}{(n-k+1)(n-k)!(k-1)!} = \\
&= \frac{(n-k+1)n!}{k(n-k+1)(n-k)!(k-1)!} + \frac{kn!}{k(n-k+1)(n-k)!(k-1)!} = \\
&= \frac{(n-k+1)n! + kn!}{k(n-k+1)(n-k)!(k-1)!} = \frac{n!((n-k+1)+k)}{k(n-k+1)(n-k)!(k-1)!} = \\
&= \frac{n!(n+1)}{k(n-k+1)(n-k)!(k-1)!} = \frac{(n+1)!}{((n+1)-k)!k!} = C(n+1,k)
\end{aligned}$$

as desired.

1.8.15

(a) Prove that

$$\sum_{i=0}^{n} C(n,i) = 2^n$$

We can follow that from the fact that there are 2^n subsets of any given finite set, which means that the number of subsets of different lengths sums up to 2^n.

Another way to do this is to use binomial theorem:

$$(x+y)^n = \sum_{i=0}^{n} C(n,i)x^k y^{n-k}$$

thus if we subisitute x and y for 1, we get

$$(1+1)^n = \sum_{i=0}^{n} C(n,i)1^k 1^{n-k}$$

$$2^n = \sum_{i=0}^{n} C(n,i)$$

(b) Prove that

$$\sum_{i=0}^{n} (-1)^i C(n,i) = 0$$

I'm sure that there is a neat explanation for this one as well, but using the binomial theorem once again, but now substituting 1 for x and -1 for y we get

$$(1-1)^n = \sum_{i=0}^{n} C(n,i)1^i(-1)^{n-i}$$

$$\sum_{i=0}^{n} C(n,i)1^i(-1)^{n-i} = 0$$

we can follow through the even-odd argument that $1^i(-1)^{n-i} = (-1)^i$, but I'll skip it.

1.8.19

(rewording) Prove the formula for unordered sampling with replacement.

This thing is ought to be covered rigorously in a course for discrete maths, combinatorics or something of sorts. Currentry there is a better proof at Belcastro's "Discrete mathematics with ducks".

1.8.20

Prove the binomial theorem 1.8.2

1.8.2 states that

$$(x+y)^n = \sum_{i=0}^{n} C(n,i)x^i y^{n-i}$$

Let

$$I = \{n \in \omega : (x+y)^n = \sum_{i=0}^{n} C(n,i)x^i y^{n-i}\}$$

We follow that

$$(x+y)^0 = C(0,0)x^0y^0 = 1$$

Thus $0 \in I$. (we can start with a base case of 1 as well for a more clear example, but I like this one more, and it suffices as well).

Now suppose that $n \in I$. We follow that

$$(x+y)^n = \sum_{i=0}^{n} C(n,i)x^i y^{n-i}$$

thus we follow that

$$(x+y)(x+y)^n = (x+y)\left[\sum_{i=0}^{n} C(n,i)x^i y^{n-i}\right]$$

Left-hand side is reduced to

$$(x+y)(x+y)^n = (x+y)^{n+1}$$

Right-hand side is obviously a bit trickier, but we can follow

$$(x+y)\sum_{i=0}^{n} C(n,i)x^i y^{n-i} =$$

$$= x\sum_{i=0}^{n} C(n,i)x^i y^{n-i} + y\sum_{i=0}^{n} C(n,i)x^i y^{n-i} =$$

$$= \sum_{i=0}^{n} C(n,i)x^{i+1} y^{n-i} + \sum_{i=0}^{n} C(n,i)x^i y^{n+1-i} =$$

$$= \sum_{i=0}^{n} C(n,i)x^i y^{n+1-i} + \sum_{i=0}^{n} C(n,i)x^{i+1} y^{n-i} =$$

$$= C(n,n)x^{n+1}y^0 + \sum_{i=0}^{n} C(n,i)x^i y^{n+1-i} + \sum_{i=0}^{n-1} C(n,i)x^{i+1} y^{n-i} =$$

$$= x^{n+1} + \sum_{i=0}^{n} C(n,i)x^i y^{n+1-i} + \sum_{i=0}^{n-1} C(n,i)x^{i+1} y^{n-i} =$$

$$= x^{n+1} + \sum_{i=0}^{n} C(n,i)x^i y^{n+1-i} + x\sum_{i=0}^{n-1} C(n,i)x^i y^{n-i} =$$

$$= x^{n+1} + \sum_{i=0}^{n} C(n,i)x^i y^{n+1-i} + x\sum_{i=1}^{n} C(n,i-1)x^{i-1} y^{n-(i-1)} =$$

$$= x^{n+1} + C(n,0)x^0 y^{n+1} + \sum_{i=1}^{n} C(n,i)x^i y^{n+1-i} + \sum_{i=1}^{n} C(n,i-1)x^i y^{n+1-i} =$$

$$= x^{n+1} + y^{n+1} + \sum_{i=1}^{n} C(n,i)x^i y^{n+1-i} + \sum_{i=1}^{n} C(n,i-1)x^i y^{n+1-i} =$$

$$= x^{n+1} + y^{n+1} + \sum_{i=1}^{n} (C(n,i) + C(n,i-1))x^i y^{n+1-i} =$$

$$= x^{n+1} + y^{n+1} + \sum_{i=1}^{n} C(n+1,i)x^i y^{n+1-i} = x^{n+1} + C(n+1,0)x^0 y^{n+1-0} + \sum_{i=1}^{n} C(n+1,i)x^i y^{n+1-i} =$$

$$= x^{n+1} + \sum_{i=0}^{n} C(n+1,i)x^i y^{n+1-i} = x^{n+1}y^0 + \sum_{i=0}^{n} C(n+1,i)x^i y^{n+1-i} =$$

$$= C(n+1,n+1)x^{n+1} y^{n+1-(n+1)} + \sum_{i=0}^{n} C(n+1,i)x^i y^{n+1-i} = \sum_{i=0}^{n+1} C(n+1,i)x^i y^{n+1-i}$$

Thus we follow

$$(x+y)^{n+1} = \sum_{i=0}^{n+1} C(n+1,i)x^i y^{n+1-i}$$

or

$$(x+y)^{n^+} = \sum_{i=0}^{n^+} C(n^+,i)x^i y^{n^+-i}$$

which means that $n \in I \Rightarrow n^+ \in I$, from which we conclude that $I = \omega$, and thus

$$(x+y)^n = \sum_{i=0}^{n} C(n,i)x^i y^{n-i}$$

for all $n \in \omega$, as desired.

1.8.22

Skip

1.9 Multinomial Coefficients

1	$(21!)/(7!*7!*7!)$
2	$50!/(18!*12!*12!*8!)$
3	$300!/(5!*8!*287!)$
4	$(3!3!2!)/10! = 1/50400$
5	$M(n,(n_1,...,n_6))/6^n$
6	$(7!)/(2*6^7)$
7	$M(12,(6,2,4))*M(13,(4,6,3))/M(25,(10,8,7))$
8	$M(12,(3,3,3,3)*M(40,(10,10,10,10))/M(52,(13,13,13,13)$
9	$4!/M(52,(13,13,13,13)$
10	$(2!*3!*4!)/9!$

1.10 The Probability of a Union of Events

1	≈ 0.11913
2	85
3	45

1.10.1

$$Pr(A_1) = Pr(A_2) = Pr(A_3) = C(4,2)*C(48,3)/C(52,5)$$

$$Pr(A_1 \cup A_2) = Pr(A_1 \cup A_3) = Pr(A_2 \cup A_3) = C(4,2)*C(48,3)*C(45,3)/C(52,5)^2$$

$$Pr(A_1 \cup A_2 \cup A_3) = 0$$

$$Pr(A_1 \cup A_2 \cup A_3) = 3*C(4,2)*C(49,3)/C(52,5) - 3C(4,2)*C(49,3)*C(46,3)/C(52,5)^2$$

TODO later (probably never).

Chapter 2

Conditional Probability

2.1 Definition of Conditional Probability

1	$Pr(A)/Pr(B)$
2	0
3	$Pr(A)$
4	$1/27 \approx 0.037037$
5	-
6	2/3
7	1/3
8	$0.6/0.85 \approx 0.706$
9a	3/4
9b	3/5
10	0.4485884485884486
11	-
12	-
13	4/9
14	0.056
15	0.47
16	5/12
17	-

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

2.1.5

$$\frac{r}{r+b} * \frac{(r+k)}{(r+k)+b} * \frac{(r+2k)}{(r+2k)+b} * \frac{b}{(r+3k)+b}$$

2.1.6

Let A be an event, that we've picked up a card, looked at its side and that the side is green. We can follow that

$$Pr(A) = 1/2$$

Let B be an event that we've picked up a card, and it's green on both sides. We follow that

$$Pr(B) = 1/3$$

Probability that both A and B happened are 1/3. Thus we follow that

$$Pr(B|A) = \frac{Pr(A \cap B)}{Pr(A)} = \frac{1/3}{1/2} = 2/3$$

This makes me think about Monty Hall problem, as those two are (probably) closely related.

2.1.11

We want to prove that

$$Pr(A^c|B) = 1 - Pr(A|B)$$

we follow that by

$$Pr(A^c|B) = \frac{Pr(A^c \cap B)}{Pr(B)} = \frac{Pr(B) - Pr(A \cap B)}{Pr(B)} = 1 - \frac{Pr(A \cap B)}{Pr(B)} = 1 - Pr(A|B)$$

where

$$Pr(A^c \cap B) = Pr(B) - Pr(A \cap B)$$

is proven in Theorem 1.5.6. as desired.

2.1.12

$$Pr(A \cup B|D) = \frac{Pr((A \cup B) \cap D)}{Pr(D)} = \frac{Pr((A \cap D) \cup (B \cap D))}{Pr(D)} =$$

$$= \frac{Pr(A \cap D) + Pr(B \cap D) - Pr(A \cap D \cap B \cap D)}{Pr(D)} =$$

$$= \frac{Pr(A \cap D) + Pr(B \cap D) - Pr(A \cap B \cap D)}{Pr(D)} =$$

$$= \frac{Pr(A \cap D)}{Pr(D)} + \frac{Pr(B \cap D)}{Pr(D)} - \frac{Pr(A \cap B \cap D)}{Pr(D)} = Pr(A|D) + Pr(B|D) - Pr(A \cap B|D)$$

every deriviation that was done here was either justified by a theorem in section 1.5 or is a property of set operations.

2.1.17

We can't have

$$Pr((A|C)|B)$$

on the account that $A|C$ is not an event, but just a funky notation introduced with the probability function. What this notation gives is just a syntactic sugar.

$$Pr(A|C) = \frac{Pr(A \cap C)}{Pr(C)} = \frac{1}{Pr(C)} Pr(A \cap C) = \frac{1}{Pr(C)} \sum_{j=1}^{n} Pr(B_j) Pr(A \cap C|B_j) =$$

$$= \frac{1}{Pr(C)} \sum_{j=1}^{n} Pr(B_j) \frac{Pr(A \cap C \cap B_j)}{Pr(B_j)} = \sum_{j=1}^{n} Pr(B_j) \frac{Pr(A \cap C \cap B_j)}{Pr(B_j) Pr(C)} =$$

$$= \sum_{j=1}^{n} \frac{Pr(A \cap C \cap B_j)}{Pr(C)} = \sum_{j=1}^{n} \frac{Pr(B_j \cap C) Pr(A \cap C \cap B_j)}{Pr(B_j \cap C) Pr(C)} =$$

$$= \sum_{j=1}^{n} \frac{Pr(B_j \cap C) Pr(A \cap B_j \cap C)}{Pr(C) Pr(B_j \cap C)} =$$

$$= \sum_{j=1}^{n} \frac{Pr(B_j \cap C)}{Pr(C)} * \frac{Pr(A \cap B_j \cap C)}{Pr(B_j \cap C)} = \sum_{j=1}^{n} Pr(B_j|C) Pr(A|B_j \cap C)$$

assuming that $Pr(B_j \cap C), Pr(C) \neq 0$ for all $1 \leq j \leq n$.

2.2 Independent Events

1	$Pr(A^c)$
2	-
3	-
4	$1/216$
5	$1-10^{-6}$
6	$149/5000 = 0.0298$
7a	$23/25 = 0.92$
7b	$20/23 \approx 0.869565$
8	$1/36 \approx 0.0277778$
9	$1/7 \approx 0.142857$
10	$\frac{106}{781} \approx 0.1357234314980794$
11	$67/256 = 0.26171875$
12a	$3/4 = 0.75$
12b	$11/24 \approx 0.4583333333$
13	0.09135172474836409
14	0.09561792499119552
15	161

2.2.1

Suppose that A and B are independent events. Thus

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

and

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

thus

$$Pr(A^c|B^c) = \frac{Pr(A^c \cap B^c)}{Pr(B^c)} = \frac{Pr((A \cup B)^c)}{Pr(B^c)} = \frac{1 - Pr(A \cup B)}{Pr(B^c)} =$$

$$= \frac{1 - (Pr(A) + Pr(B) - Pr(A)Pr(B))}{Pr(B^c)} = \frac{1 - Pr(A) - Pr(B) + Pr(A)Pr(B))}{Pr(B^c)} =$$

$$= \frac{1 - Pr(B) - Pr(A) + Pr(A)Pr(B))}{Pr(B^c)} = \frac{1 - Pr(B)}{Pr(B^c)} + \frac{-Pr(A) + Pr(A)Pr(B))}{Pr(B^c)} =$$

$$= 1 + \frac{Pr(A)(-1 + Pr(B))}{Pr(B^c)} = 1 - \frac{Pr(A)(1 - Pr(B))}{Pr(B^c)} = 1 - Pr(A)\frac{1 - Pr(B)}{Pr(B^c)} =$$

$$= 1 - Pr(A) = Pr(A^c)$$

Same goes for $Pr(B^c|A^c)$

2.2.2

2.2.1 implies that

$$Pr(A^c) = Pr(A^c|B^c)$$

and

$$Pr(B^c) = Pr(B^c|A^c)$$

for the nonzero cases, and if $Pr(A) = 0$ or $Pr(B) = 0$, then the cases are trivial.

2.2.3

Suppose that A is an event and $Pr(A) = 0$ and B is another event. We follow that

$$Pr(A \cap B) \leq Pr(A)$$

and thus

$$Pr(A \cap B) = 0$$

as desired.

2.2.7b

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

2.2.9

Assuming $1 \leq n \leq \infty$

$$\sum (p_n)^3 = \sum (2^{-n})^3 = \sum 2^{-3n} = \sum (1/8)^n = \frac{1/8}{1 - 1/8} = 1/7$$

2.2.10

Let A be an event that at least 1 child in the family has blue eyes and let B be an event that at least 3 children have blue eyes. We follow that

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

given that $B \subseteq A$, we follow that

$$Pr(B|A) = \frac{Pr(B)}{Pr(A)}$$

We follow that

$$Pr(A) = 1 - (1 - 1/4)^5 = 781/1024$$

and

$$Pr(B) = \sum_{i\in\{3,4,5\}} C(n,i)1/4 * C(n,n-i)(1-1/4) = \sum_{i\in\{3,4,5\}} C(n,i)(1/4)^i(3/4)^{5-i} = 53/512$$

thus

$$Pr(B|A) = \frac{Pr(B)}{Pr(A)} = \frac{106}{781} \approx 0.1357234314980794$$

2.2.11

If the youngest child in the family has the blue eyes, then we can't say that $B \subseteq A$. Given that the probabilitiees of children having different colored eyes are independent, we follow that we can rewrite this problem as "what's the probability of that the remaining 4 children have at least 2 blue-eyed children among them". This happens to be equal to

$$\sum_{i\in\{2,3,4\}} C(4,i)(1/4)^i(3/4)^{4-i} = 67/256 = 0.26171875$$

Done with this section; moving on

2.3 Bayes' Theorem

1	-
2	3
3	0.3
4	0.0001899658061548921
5	0.30508474576271183
6a	0.9896907216494846
6b	0.9846153846153847
7a	0, 1/10, 1/5, 3/10, 2/5
8	skip
16	-

2.3.1

Suppose that S can be partitioned into $B_1, ..., B_k$. Suppose also that A is an event such that $Pr(A) > 0$ and

$$Pr(B_1|A) < Pr(B_1)$$

and

$$Pr(B_i|A) \leq Pr(B_i)$$

for all $1 < i \leq k$. Thus we follow that

$$\sum Pr(B_i|A) < \sum Pr(B_i) = 1$$

thus

$$\sum Pr(B_i|A) < 1$$

$$\sum \frac{Pr(B_i \cap A)}{Pr(A)} < 1$$

$$\sum Pr(B_i \cap A) < Pr(A)$$

Given that B_i is a partition of S, we follow that B_i's are disjoint (BTW if several sets are all pairwise disjoint, then all of them are disjoint), therefore we follow that $B_j \cap A$ is disjoint from $B_l \cap A$ for all $1 \leq j, l \leq k$. Thus

$$\sum Pr(B_i \cap A) = Pr(\bigcup [B_i \cap A]) = Pr(\bigcup [B_i] \cap A) = Pr(S \cap A) = Pr(A) < Pr(A)$$

which is a contradiction.

2.3.16

(a)

Suppose that D_1 is independent of B. That is,

$$Pr(D_1) = Pr(D_1|B) = 0.01$$

Assume that for some n we've got that

$$Pr(D_n) = 0.01$$

We follow that

$$Pr(D_{n+1}|B) = 0.01$$

If B^c is true and we know that n'th item is normal, then we can follow that

$$Pr(D_{n+1}|D_n^c \cap B^c) = 1/165$$

If n'th item is defective, then

$$Pr(D_{n+1}|D_n \cap B^c) = 2/5$$

therefore, because D and D^c are partitioning space, we follow that

$$Pr(D_{n+1}|B^c) = Pr(D_n^c) * 1/165 + Pr(D_n) * 2/5 = 0.01$$

thus we now can follow that

$$Pr(D_{n+1}) = 0.1 * 0.7 + 0.01 * 0.3 = 0.1$$

therefore by induction we can conclude that $Pr(D_n) = 0.01$ for all $n \in N$

(b)

Let us assume that we've got a typo in the text, and we actually need to compute $Pr(B|E)$. From our initial assumptions we follow that

$$Pr(E|B) = 0.99^4 * 0.01^2 = 9.65 * 10^{-5}$$

thus we need to compute

$$Pr(B|E) = \frac{Pr(E|B) * Pr(B)}{Pr(E|B) * Pr(B) + Pr(E|B^c) * Pr(B^c)}$$

thus the only thing that we need to compute is $Pr(E|B^c)$. We follow that

$$Pr(E|B^c) =$$

$$= Pr(D_1^c \cap D_2^c \cap D_3 \cap D_4 \cap D_5^c \cap D_6^c|B^c) = Pr(D_1^c|B^c)Pr(D_2^c|D_1^c \cap B)Pr(D_3|D_2^c \cap B)... =$$

$$= 0.99 * 164/165 * 1/165 * 2/5 * 3/5 * 164/165 = 0.99 * (164/165)^2 * 1/165 * 2/5 * 3/5 =$$

$$= 0.001422598347107438$$

thus we can now compute the rest and state that

$$Pr(B|E) = 0.11898006688921978 \approx 12\%$$

2.4 The Gambler's Ruin Promlem

1	-
2	all the same
3	a
4	c
5	198
6	7
7	-

2.4.1

Suppose that we've got conditions from Example 2.4.2. Let i be a natural number such that $i \leq 98$. Probability that gambler A's gonna win i dollars before losing $100 - i$ is

$$a_i = \frac{(3/2)^i - 1}{(3/2)^{100} - 1}$$

we follow that a_i is an increasing function and thus we can conclude that in order to get the desired conclusion, we need to calculate the case $i = 98$. We follow that

$$a_{98} = \frac{(3/2)^{98} - 1}{(3/2)^{100} - 1} \approx 0.444444$$

BTW, it's not a pretty rational number.

2.4.7

we follow that

$$f_i = \frac{(1/3)^i - 1}{(1/3)^{i+2} - 1}$$

is the desired function. We want to show that the function is decreasing and $a_1 < 1/4$. Simple calculation show that $a_1 \approx 0.14285714285714282$. We also follow that

$$f_n - f_{n+1} = \frac{(1/3)^n - 1}{(1/3)^{n+2} - 1} - \frac{(1/3)^{n+1} - 1}{(1/3)^{n+3} - 1}$$

Maxima shows that this thing is equal to

$$-\frac{16 * 3^{n+2}}{something.positive}$$

which is good enough for me to prove that this thing is always below 1/4, as desired.

Done with this section

Chapter 3

Random Variables and Distributions

3.1 Random Variables and Discrete Distributions

Notes

Let S be a sample space

A random variable is a function $f : S \to R$ (which is confusing). Canonical representations of random variables are capital English letters (i.e. $X, Y, etc.$)

If X is a random variable, then R can be thought of as a sample space, and then we can define $g : \mathcal{P}(R) \to R$ such that

$$g(C) = Pr(\{s \in S : X(s) \in C\})$$

Function g is then called distribution of X, and it is a probability measure on S (in the book, definition of probability is somewhat un-intuitive, but the not just below the definition produces given definition)

Random variable X has a discrete distribution (we also say that X is a discrete variable) if range(X) is countable. Distribution of the variable does not show up in the definition, which is not helpful.

In general, we abuse notation in this book a lot. It's not hard to decode what $Pr(X = c)$ is supposed to mean.

If X is discrete and g is a distribution of X, then a probability function $f : R \to R$ is a function such that $f(x) = g(\{x\})$ Closure of $\{x \in R : f(x) > 0\}$ is called a support of X. Although the term "closure" was not defined anywhere in this book, from the internet it seems that it is the topological notion of closure. It is nice that in the course of that book we've defined union, but we haven't mentioned closure at all.

If random variable X has range $\{0, 1\}$ with $Pr(X = 1) = p$ for some $p \in R$, then we say that X has the Bernoulli distribution with parameter p. Although it might seem that any given random variable with range of size of 2 has Bernoulli distribution, I'm pretty sure that we won't encounter such variables in the wild, but if for some reason I will ever have to use such a function, I'll name its distribution proto-Bernoulli.

Random variable X has uniform distribution of the integers $a, ..., b$ (i.e. $Z \cap [a, b]$ for some $a, B \in Z$ such that $a \leq b$) if $f(m) = f(n)$ for all $m, n \in Z \cap [a, b]$.

1	6/11
2	1/15
3	no
4	binomial with 10 and 1/2
5	skip
6	0.15087890625
7	0.80589565
8	0.13295332343433508
9	1/2
10a	1/120 (x + 1)(8 - x)
10b	1/3
11	harmonics

3.2 Continous Distributions

1	4/9
2	31/48, 9/16, 136/243
3	1/2, 13/27, 2/27

The rest of that damned section is just exercises in trivial calculus. Skipping all this stuff.

3.3 The Cumulative Distribution Function

Notes

c.d.f. is a really nice way to describe distribution of a given random variable. Firstly, we don't care whether or not the variable is discrete, continous, or whatever, it's got to have a c.d.f. Secondly, it's an increasing bounded function from reals to reals, which implies that any discontinuity in a given function is a jump discontinuity (see section 4.6. in real analysis course) and the set of its discontinuities is countable. Also, given distribution of a random variable, the c.d.f. is unique.

If we get into deeper parts of the book, we can conclude that we can ditch somewhat non-rigorous notions of p.d.f and p.f. and concentrate exclusively on c.d.f. of a given random variable for all the theoretical parts.

3.4 Bivariate Distributions

1	1/2, 1/4
2	0.27, ...
3	1/40, 1/20, 0.175(7/40), 7/10

Exercises are a bunch of integrals/sums and other borderline trivial stuff. This is practically an exercise in Maxima. Skip

3.5 Marginal Distributions

1	f(y) = (d - c)k, f(y) = (b - a) k
2	$f(y) = \frac{y+1}{30}, f(x) = \frac{2x+3}{15}$, dependent

this section is also an exercise in calculus.

3.6 Conditional Distributions

3.6.1

$$f(x|y) = \frac{3y^2}{2x^2 - 1^{3/2}}$$

for appropriate values of x and y.

The rest is calculus

3.7 Multivariate Distributions

1	1/3, $x_3 + 1/3x_1 + 1/3$, $5/13 \approx 0.38461538461538464$
2	6, $6\beta(4 - k - k, k + j + 2)$, TODO