Chapter 1

Probability Theory

“If any little problem comes your way, I shall be happy, if I can, to give you a hint or two as to its solution.”

Sherlock Holmes

The Adventure of the Three Students

1.1 a. Each sample point describes the result of the toss (H or T) for each of the four tosses. So,
 for example THTT denotes T on 1st, H on 2nd, T on 3rd and T on 4th. There are 24 = 16
 such sample points.

b. The number of damaged leaves is a nonnegative integer. So we might use S = {0, 1, 2, . . .}.

c. We might observe fractions of an hour. So we might use S = {t : t ≥ 0}, that is, the half
 infinite interval [0, ∞).

d. Suppose we weigh the rats in ounces. The weight must be greater than zero so we might use
 S = (0,∞). If we know no 10-day-old rat weighs more than 100 oz., we could use S = (0,100].

e. If n is the number of items in the shipment, then S = {0/n, 1/n, . . . , 1}.

1.2 For each of these equalities, you must show containment in both directions.

a. x ∈ A\B ⇔ x ∈ A and x ∈ B ⇔ x ∈ A and x ∈ A ∩ B ⇔ x ∈ A\(A ∩ B). Also, x ∈ A and
 x ∈ B ⇔ x ∈ A and x ∈ Bc ⇔ x ∈ A ∩ Bc.

b. Suppose x ∈ B. Then either x ∈ A or x ∈ Ac. If x ∈ A, then x ∈ B ∩ A, and, hence
 x ∈ (B ∩A)∪(B ∩Ac). Thus B ⊂ (B ∩A)∪(B ∩Ac). Now suppose x ∈ (B ∩A)∪(B ∩Ac).
 Then either x ∈ (B ∩ A) or x ∈ (B ∩ Ac). If x ∈ (B ∩ A), then x ∈ B. If x ∈ (B ∩ Ac),
 then x ∈ B. Thus (B ∩ A) ∪ (B ∩ Ac) ⊂ B. Since the containment goes both ways, we have
 B = (B ∩ A) ∪ (B ∩ Ac). (Note, a more straightforward argument for this part simply uses
 the Distributive Law to state that (B ∩ A) ∪ (B ∩ Ac) = B ∩ (A ∪ Ac) = B ∩ S = B.)

c. Similar to part a).

d. From part b).

A ∪ B = A ∪ [(B ∩ A) ∪ (B ∩ Ac)] = A ∪ (B ∩ A) ∪ A ∪ (B ∩ Ac) = A ∪ [A ∪ (B ∩ Ac)] = A ∪ (B ∩ Ac).

1.3 a. x ∈ A ∪ B ⇔ x ∈ A or x ∈ B ⇔ x ∈ B ∪ A

x ∈ A ∩ B ⇔ x ∈ A and x ∈ B ⇔ x ∈ B ∩ A.

b. x ∈ A ∪ (B ∪ C) ⇔ x ∈ A or x ∈ B ∪ C ⇔ x ∈ A ∪ B or x ∈ C ⇔ x ∈ (A ∪ B) ∪ C.
 (It can similarly be shown that A ∪ (B ∪ C) = (A ∪ C) ∪ B.)

x ∈ A ∩ (B ∩ C) ⇔ x ∈ A and x ∈ B and x ∈ C ⇔ x ∈ (A ∩ B) ∩ C.

c. x ∈ (A ∪ B)c ⇔ x ∈ A or x ∈ B ⇔ x ∈ Ac and x ∈ Bc ⇔ x ∈ Ac ∩ Bc

x ∈ (A ∩ B)c ⇔ x ∈ A ∩ B ⇔ x ∈ A and x ∈ B ⇔ x ∈ Ac or x ∈ Bc ⇔ x ∈ Ac ∪ Bc.

1.4 a. “A or B or both” is A∪B. From Theorem 1.2.9b we have P (A∪B) = P (A)+P (B)−P (A∩B).

1-2 Solutions Manual for Statistical Inference

b. “A or B but not both” is (A ∩ Bc) ∪ (B ∩ Ac). Thus we have

P ((A ∩ Bc) ∪ (B ∩ Ac)) = P(A ∩ Bc) + P(B ∩ Ac) (disjoint union)

= [P (A) − P (A ∩ B)] + [P (B) − P (A ∩ B)] (Theorem1.2.9a)

= P (A) + P (B) − 2P (A ∩ B).

c. “At least one of A or B” is A ∪ B. So we get the same answer as in a).

d. “At most one of A or B” is (A ∩ B)c, and P ((A ∩ B)c) = 1 − P (A ∩ B).

1.5 a. A ∩ B ∩ C = {a U.S. birth results in identical twins that are female}

b. P (A ∩ B ∩ C) =~~1~~90 ×3 ×2

1.6

p0 = (1 − u)(1 − w), p1 = u(1 − w) + w(1 − u), p2 = uw,

p0 = p2 ⇒ u+w=1

p1 = p2 ⇒ uw = 1/3.

These two equations imply u(1 − u) = 1/3, which has no solution in the real numbers. Thus, the probability assignment is not legitimate.

1.7 a.

{

1−~~πr2~~ if i = 0

P (scoring i points) =

b.

[A ]

πr2 (6−i)2 −(5−i)2

A 52

if i = 1, . . . , 5.

P (scoring i points|board is hit)

P (board is hit)

= ~~P~~~~(scoringipoints∩boardishit)~~
 P (board is hit)

= ~~π~~~~r2~~

A

P (scoring i points ∩ board is hit)

[ (6 − i)2 − (5 − i)2 = ~~π~~~~r2~~

]
 i = 1,...,5.

Therefore,

P (scoring i points|board is hit) =

A 52

(6 − i)2 − (5 − i)2 52

i = 1,...,5

which is exactly the probability distribution of Example 1.2.7.

1.8 a. P (scoring exactly i points) = P (inside circle i) − P (inside circle i + 1). Circle i has radius
 (6 − i)r/5, so

P (sscoring exactly i points) =

π(6 − i)2r2 52πr2

((6−(i + 1)))2r2

− π = ~~(~~~~6−i)2~~~~−~~~~(5−i)2~~

52πr2 52

11−2i

b. Expanding the squares in part a) we find P (scoring exactly i points) =
 decreasing in i.

25

, which is

c. Let P (i) =~~11−2i~~. Since i ≤ 5, P (i) ≥ 0 for all i. P (S) = P (hitting the dartboard) = 1 by

25

definition. Lastly, P (i ∪ j) = area of i ring + area of j ring = P (i) + P (j).

1.9 a. Suppose x ∈ (∪αAα)c, by the definition of complement x ∈ ∪αAα, that is x ∈ Aα for all
 α ∈ Γ. Therefore x ∈ Acα forallα∈Γ.Thusx∈∩αAα and,bythedefinitionofintersection
 x ∈ Acα forallα∈Γ.Bythedefinitionofcomplementx∈Aα forallα∈Γ.Therefore
 x ∈ ∪αAα. Thus x ∈ (∪αAα)c.

Second Edition 1-3

b. Suppose x ∈ (∩αAα)c, by the definition of complement x ∈ (∩αAα). Therefore x ∈ Aα for
 some α ∈ Γ. Therefore x ∈ Acα forsomeα∈Γ.Thusx∈∪αAα and,bythedefinitionof
 union, x ∈ Acα forsomeα∈Γ.Thereforex∈Aα forsomeα∈Γ.Thereforex∈∩αAα.Thus
 x ∈ (∩αAα)c.

1.10 For A1, . . . , An

(

(i)

⋃

i=1

)c

⋂

Ai =

i=1

Ac

i

( )c

⋂ ⋃

(ii) Ai = Ac

i

i=1 i=1

Proof of (i): If x ∈ (∪Ai)c, then x ∈ ∪Ai. That implies x ∈ Ai for any i, so x ∈ Aci foreveryi

and x ∈ ∩Ai.

Proof of (ii): If x ∈ (∩Ai)c, then x ∈ ∩Ai. That implies x ∈ Aci forsomei,sox∈∪Ai.

1.11 We must verify each of the three properties in Definition 1.2.1.

a. (1) The empty set ∅ ∈ {∅, S}. Thus ∅ ∈ B. (2) ∅c = S ∈ B and Sc = ∅ ∈ B. (3) ∅∪S = S ∈ B.

b. (1) The empty set ∅ is a subset of any set, in particular, ∅ ⊂ S. Thus ∅ ∈ B. (2) If A ∈ B,

then A ⊂ S. By the definition of complementation, Ac is also a subset of S, and, hence,
Ac ∈ B. (3) If A1,A2,... ∈ B, then, for each i,Ai ⊂ S. By the definition of union, ∪Ai ⊂ S.
Hence, ∪Ai ∈ B.

c. Let B1 and B2 be the two sigma algebras. (1) ∅ ∈ B1 and ∅ ∈ B2 since B1 and B2 are

sigma algebras. Thus ∅ ∈ B1 ∩ B2. (2) If A ∈ B1 ∩ B2, then A ∈ B1 and A ∈ B2. Since

B1 and B2 are both sigma algebra Ac ∈ B1 and Ac ∈ B2. Therefore Ac ∈ B1 ∩ B2. (3) If

A1,A2,... ∈ B1 ∩ B2, then A1,A2,... ∈ B1 and A1,A2,... ∈ B2. Therefore, since B1 and B2

are both sigma algebra, ∪∞i=1Ai ∈B1 and∪

1.12 First write
 ( ) (

=1Ai ∈B2.Thus∪ =1Ai ∈B1 ∩B2.

)

⋃

P Ai = P

i=1

⋃

Ai ∪

i=1

( )

⋃

Ai

i=n+1

( )

⋃

= P Ai

i=1

∑

+P

(

⋃

Ai (Ais are disjoint)

i=n+1

)

⋃

= P (Ai) + P Ai (finite additivity)

i=1 i=n+1

Now define Bk =⋃∞i=k Ai.NotethatBk+1 ⊂Bk andBk →φask→∞.(Otherwisethesum

of the probabilities would be infinite.) Thus

( ) ( )

[ ]

⋃ ⋃ ∑ ∑

P Ai = lim P Ai = lim P (Ai) + P (Bn+1) = P (Ai).

n→∞ n→∞

i=1 i=1 i=1 i=1

1.13 If A and B are disjoint, P (A ∪ B) = P (A) + P (B) =13 +4 =~~12~~, which is impossible. More

generally, if A and B are disjoint, then A ⊂ Bc and P (A) ≤ P (Bc). But here P (A) > P (Bc),

so A and B cannot be disjoint.

1.14 If S = {s1, . . . , sn}, then any subset of S can be constructed by either including or excluding
 si, for each i. Thus there are 2n possible choices.

1.15 Proof by induction. The proof for k = 2 is given after Theorem 1.2.14. Assume true for k, that
 is, the entire job can be done in n1 × n2 × · · · × nk ways. For k + 1, the k + 1th task can be
 done in nk+1 ways, and for each one of these ways we can complete the job by performing

n

n

n

n

i

i

∞

n

∞

n

∞

n

∞

∞

∞

n

∞

1-4 Solutions Manual for Statistical Inference

the remaining k tasks. Thus for each of the nk+1 we have n1 × n2 × · · · × nk ways of com-

pleting the job by the induction hypothesis. Thus, the number of ways we can do the job is

(1 × (n1 × n2 × · · · × nk )) + · · · + (1 × (n1 × n2 × · · · × nk )) =n1 ×n2 ×···×nk ×nk+1.

| {z }

nk+1terms

1.16 a) 263. b) 263 + 262. c) 264 + 263 + 262.

(n)

1.17 There are

2

= n(n − 1)/2 pieces on which the two numbers do not match. (Choose 2 out of

n numbers without replacement.) There are n pieces on which the two numbers match. So the total number of different pieces is n + n(n − 1)/2 = n(n + 1)/2.

1.18 The probability is

(n2)n! nn

= ~~(~~~~n−1)(n−1)!~~ . There are many ways to obtain this. Here is one. The

2nn−2

denominator is nn because this is the number of ways to place n balls in n cells. The numerator
is the number of ways of placing the balls such that exactly one cell is empty. There are n ways
to specify the empty cell. There are n − 1 ways of choosing the cell with two balls. There are( )

n

2

ways of picking the 2 balls to go into this cell. And there are (n − 2)! ways of placing the

remaining n − 2 balls into the n − 2 cells, one ball in each cell. The product of these is the(n) (n)

numerator n(n − 1)
 (6)

1.19 a. = 15.

2

(n − 2)! = n!.

2

4

b. Think of the n variables as n bins. Differentiating with respect to one of the variables is
 equivalent to putting a ball in the bin. Thus there are r unlabeled balls to be placed in n(n+r−1)

unlabeled bins, and there are

r

ways to do this.

1.20 A sample point specifies on which day (1 through 7) each of the 12 calls happens. Thus there

are 712 equally likely sample points. There are several different ways that the calls might be

assigned so that there is at least one call each day. There might be 6 calls one day and 1 call

each of the other days. Denote this by 6111111. The number of sample points with this pattern(12) (12)

is 7

6

6!. There are 7 ways to specify the day with 6 calls. There are to specify which of

6

the 12 calls are on this day. And there are 6! ways of assigning the remaining 6 calls to the

remaining 6 days. We will now count another pattern. There might be 4 calls on one day, 2 calls

on each of two days, and 1 call on each of the remaining four days. Denote this by 4221111.(12)(6)(8)(6)

The number of sample points with this pattern is 7

4!. (7 ways to pick day with 4

calls,

(12)

4

(6)

to pick the calls for that day,
 (6) 2

4 2 2 2 (8)

to pick two days with two calls, ways to pick

2

two calls for lowered numbered day,

2

ways to pick the two calls for higher numbered day,

4! ways to order remaining 4 calls.) Here is a list of all the possibilities and the counts of the sample points for each one.

pattern number of sample points

(12)

6111111 7 6! = 4,656,960

(16 ) (7)

2

5211111 7 6 5! = 83,825,280

(1

5

)(6)(8)(6)

2

4221111 7 4! = 523,908,000

(14 ) (8)2 2

2

4311111 7 6 5! = 139,708,800

(7)(12

4

)( )

(6)

9

3321111 5

4! = 698,544,000

(12)(6)(9)(2)(5)

7

3222111 7

3! = 1,397,088,000

(7)(123 )(103)(82)(62)(4)

2222211

5 2 2 2 2 2

2! = 314,344,800

3,162,075,840

The probability is the total number of sample points divided by 712, which is~~3,162,075,840~~ ≈

712

.2285.

(n (2n)

1.21 The probability is
 (2n)

2r)22r

(2n2r )

. There are ways of choosing 2r shoes from a total of 2n shoes.

2r

Thus there are

2r

equally likely sample points. The numerator is the number of sample points
 (n)

for which there will be no matching pair. There are

2r

ways of choosing 2r different shoes

2

2

3

Second Edition 1-5

styles. There are two ways of choosing within a given shoe style (left shoe or right shoe), which(n)

gives 22r ways of arranging each one of the

arrays. The product of this is the numerator

(n) 2r

2r

1.22 a)

1.23

22r.

(3115)(2915)(3115)(30 15)···(15)

(366180)

b)~~366365~~ ~~·~~~~··336~~

(36630 )

∑

P( same number of heads ) =

P (1st tosses x, 2nd tosses x)

x=0

)x ( )n−x]2 )nn )2

=

∑[(n)(1 1

x 2 2

(1 ∑(n
=

4 x

1.24 a.

∑

P (A wins) =

i=1

x=0 x=0

P (A wins on ith toss)
 )2 )4 ( ) )2i+1

=

1 (1 1 (1 1

2+ 2 2+ 2 2

∑(1

+··· = = 2/3.

2

i=0

p

b. P (A wins) = p + (1 − p)2p + (1 − p)4p + · · · =∑∞i=0 p(1−p)2i = 1−(1−p)2 .

( )

p

c.~~d~~

dp 1−(1−p)2

= p2 0. Thus the probability is increasing in p, and the minimum

[1−(1−p)2 ]2 >
 p

is at zero. Using L’Hôpital’s rule we find limp→0

1−(1−p)2 =

1/2.

1.25 Enumerating the sample space gives S′ = {(B, B), (B, G), (G, B), (G, G)} ,with each outcome
 equally likely. Thus P (at least one boy) = 3/4 and P (both are boys) = 1/4, therefore

P( both are boys | at least one boy ) = 1/3.

An ambiguity may arise if order is not acknowledged, the space is S′ = {(B, B), (B, G), (G, G)}, with each outcome equally likely.

1.27 a. For n odd the proof is straightforward. There are an even number of terms in the sum(n) )

( n

(0, 1, · · · , n), and

k

and , which are equal, have opposite signs. Thus, all pairs cancel

n−k

and the sum is zero. If n is even, use the following identity, which is the basis of Pascal’s(n) (n−1) (n−1)

triangle: For k > 0,
 ∑

k

= + . Then, for n even

k k−1

) ) ∑ ) )

(−1)k

k=0

(n

k

(n

= +

0

)

(n

(n (n

(−1)k +

k n

k=1

) ) )]

(n ∑ [(n − 1 (n − 1

=

0

+ + (−1)k +

n k k−1

k=1

=

)

(n

0

) ( ) )

(n n−1 (n − 1

+ − − = 0.

n 0 n−1

(n) (n−1)

b. Use the fact that for k > 0, k =n to write

∑

k

)

(n

k = n

k−1

∑ ) ∑ )

(n − 1 (n − 1

=n = n2n−1.

k k−1 j

k=1 k=1 j=0

n

n

∞

∞

n

n−1

n−1

n

n

n−1

1-6

c.

Solutions Manual for Statistical Inference

∑n ∑n

k=1 (−1)k+1k(nk)= k=1(−1)k+1(n−1k)=n∑ =0 (−1)j (n−1j)=0fromparta).

1.28 The average of the two integrals is

[(n log n − n) + ((n + 1) log (n + 1) − n)] /2 = [n log n + (n + 1) log (n + 1)] /2 − n

≈ (n + 1/2) log n − n.

Let dn = log n! − [(n + 1/2) log n − n], and we want to show that limn→∞ mdn = c, a constant.

This would complete the problem, since the desired limit is the exponential of this one. This

is accomplished in an indirect way, by working with differences, which avoids dealing with the

factorial. Note that (

dn − dn+1 = n+

1

2

) ( )

1

log 1+ − 1.

n

Differentiation will show that ((n +12))log((1+n))isincreasinginn,andhasminimum value (3/2) log 2 = 1.04 at n = 1. Thus dn − dn+1 > 0. Next recall the Taylor expansion of log(1 + x) = x − x2/2 + x3/3 − x4/4 + · · ·. The first three terms provide an upper bound on log(1 + x), as the remaining adjacent pairs are negative. Hence

(
0<dndn+1 < n+

1)(1 1
2 n 2n2

)

1 1 1

+ −1= +

3n3 12n2 6n3

∑∞

It therefore follows, by the comparison test, that the series

1

dn −dn+1 converges. Moreover,

the partial sums must approach a limit. Hence, since the sum telescopes,

∑

lim

N →∞

1

dn − dn+1 = lim d1 − dN+1 = c.

N →∞

Thus limn→∞ dn = d1 − c, a constant.

Unordered Ordered

1.29 a. {4,4,12,12} (4,4,12,12), (4,12,12,4), (4,12,4,12)

(12,4,12,4), (12,4,4,12), (12,12,4,4)

Unordered Ordered

(2,9,9,12), (2,9,12,9), (2,12,9,9), (9,2,9,12) {2,9,9,12} (9,2,12,9), (9,9,2,12), (9,9,12,2), (9,12,2,9)
 (9,12,9,2), (12,2,9,9), (12,9,2,9), (12,9,9,2)

b. Same as (a).

c. There are 66 ordered samples with replacement from {1, 2, 7, 8, 14, 20}. The number of or-

dered samples that would result in {2, 7, 7, 8, 14, 14} is
Thus the probability is~~180~~ 66 .

6!

2!2!1!1!

= 180 (See Example 1.2.20).

d. If the k objects were distinguishable then there would be k! possible ordered arrangements.
 Since we have k1, . . . , km different groups of indistinguishable objects, once the positions of
 the objects are fixed in the ordered arrangement permutations within objects of the same

group won’t change the ordered arrangement. There are k1!k2! · · · km! of such permutations

for each ordered component. Thus there would be

k!

ifferent ordered components.

k1 !k2 !···km ! d

e. Think of the m distinct numbers as m bins. Selecting a sample of size k, with replacement,(k+m−1)

is the same as putting k balls in the m bins. This is

k

, which is the number of distinct

bootstrap samples. Note that, to create all of the bootstrap samples, we do not need to know
what the original sample was. We only need to know the sample size and the distinct values.

1.31 a. The number of ordered samples drawn with replacement from the set {x1, . . . , xn} is nn. The
 number of ordered samples that make up the unordered sample {x1, . . . , xn} is n!. Therefore

the outcome with average~~x1~~ ~~+~~~~x2~~ ~~+~~~~···+xn~~

that is obtained by the unordered sample {x1, . . . , xn}

n

j−1

N

Second Edition 1-7

has probability~~n!~~nn .Anyotherunorderedoutcomefrom{x1, ...,xn}, distinct from the unordered sample {x1, . . . , xn}, will contain m different numbers repeated k1, . . . , km times where k1 + k2 + · · · + km = n with at least one of the ki’s satisfying 2 ≤ ki ≤ n. The probability of obtaining the corresponding average of such outcome is

n!
k1!k2!···km!nn

< ~~n~~~~!~~ , since k1!k2! · · · km! > 1.

nn

Therefore the outcome with average~~x1~~ ~~+~~~~x2~~ ~~+~~~~···+xn~~ is the most likely.

n

b. Stirling’s approximation is that, as n → ∞, n! ≈
 )∕( √ )

√

2πnn+(1/2)e−n, and thus
√

( n! 2nπ

nn en

=  ~~√~~~~e~~~~n~~ =

nn 2nπ

2πnn+(1/2)e−nen √ = 1.

nn 2nπ

c. Since we are drawing with replacement from the set {x1, . . . , xn}, the probability of choosing

any xi is1n. Therefore the probability of obtaining an ordered sample of size n without xi

is (1 −1n)n.Toprovethatlimn→∞(1−n)n =
 ( )

e−1, calculate the limit of the log. That is
 ( )

1 log 1−1

lim n log 1− = lim

n→∞ n n→∞ 1/n

n

L’Hôpital’s rule shows that the limit is −1, establishing the result. See also Lemma 2.3.14.

1.32 This is most easily seen by doing each possibility. Let P (i) = probability that the candidate

hired on the ith trial is best. Then

P (1) =

1.33 Using Bayes rule

1

,

N

1

P (2) = ,P(i) =

N−1,...

P (CB|M )P (M )

1

,P(N) = 1.

N−i+1,...

.05 ×1

2

P (M |CB) =

1.34 a.

P (CB|M )P (M ) + P (CB|F )P (F )

= .9524.

.05 ×12+.0025×2 =

P (Brown Hair)

= P(Brown Hair|Litter 1)P(Litter 1) + P(Brown Hair|Litter 2)P(Litter 2)

) )

=

(2)(1

3 2

(3)(1 19

+ =

5 2 30.

b. Use Bayes Theorem

(2)(1)

P (Litter 1|Brown Hair) =

P (BH|L1)P (L1)
P (BH|L1)P (L1) + P (BH|L2)P (L2

3

=

2

19

30

10

=

19.

1.35 Clearly P (·|B) ≥ 0, and P (S|B) = 1. If A1, A2, . . . are disjoint, then

(



⋃ 

) ⋃∞

(⋃∞ (

P

i=1



Ai

B

i=1 Ai ∩B) = P

P (B)

∑∞

i=1 P(Ai ∩B)

i=1 (Ai ∩B))
= P

P (B)

∑

=

P (B)

= P (Ai|B).

i=1

∞

∞

1-8 Solutions Manual for Statistical Inference

1.37 a. Using the same events A, B, C and W as in Example 1.3.4, we have

P (W)

= P(W|A)P(A) + P(W|B)P(B) + P(W|C)P(C)
 ) ) )

= γ

(1

3

(1 (1

+0 +1 = ~~γ~~~~+1~~

3 3 3

Thus, P (A|W) =~~P(A∩W)~~ = γ/3

P (W) ~~(γ+1)/3~~



~~γ+1~~ where,γ 1





γ+1 = 3
 γ 1
γ+1 < 3

γ 1

if γ =1

2

if γ <1

2

γ+1 > 3 if γ >12.

b. By Exercise 1.35, P (·|W) is a probability function. A, B and C are a partition. So
 P (A|W) + P (B|W) + P (C|W) = 1.

But, P (B|W) = 0. Thus, P (A|W) + P (C|W) = 1. Since P (A|W) = 1/3, P (C|W) = 2/3. (This could be calculated directly, as in Example 1.3.4.) So if A can swap fates with C, his chance of survival becomes 2/3.

1.38 a. P (A) = P (A ∩ B) + P (A ∩ Bc) from Theorem 1.2.11a. But (A ∩ Bc) ⊂ Bc and P (Bc) =

1 − P(B) = 0. So P(A ∩ Bc) = 0, and P(A) = P(A ∩ B). Thus,

P (A|B) =

b. A ⊂ B implies A ∩ B = A. Thus,

P (A ∩ B)

P (B)

= ~~P~~~~(A)~~ = P(A)

1

P (B|A) =

And also,

P (A ∩ B)

P (A)

= ~~P~~~~(A)~~ = 1.

P (A)

P (A|B) =

P (A ∩ B)

P (B)

= ~~P~~~~(A)~~

P (B).

c. If A and B are mutually exclusive, then P (A ∪ B) = P (A) + P (B) and A ∩ (A ∪ B) = A.

Thus,

P (A|A ∪ B) =

P (A ∩ (A ∪ B))
 P (A ∪ B)

= P(A)
 P (A) + P (B).

d. P (A ∩ B ∩ C) = P (A ∩ (B ∩ C)) = P (A|B ∩ C)P (B ∩ C) = P (A|B ∩ C)P (B|C)P (C).

1.39 a. Suppose A and B are mutually exclusive. Then A ∩ B = ∅ and P (A ∩ B) = 0. If A and B
 are independent, then 0 = P (A ∩ B) = P (A)P (B). But this cannot be since P (A) > 0 and
 P (B) > 0. Thus A and B cannot be independent.

b. If A and B are independent and both have positive probability, then

0 < P(A)P(B) = P(A ∩ B).

This implies A ∩ B = ∅, that is, A and B are not mutually exclusive.

1.40 a. P (Ac ∩ B) = P (Ac|B)P (B) = [1 − P (A|B)]P (B) = [1 − P (A)]P (B) = P (Ac)P (B) , where
 the third equality follows from the independence of A and B.

b. P (Ac ∩ Bc) = P (Ac) − P (Ac ∩ B) = P (Ac) − P (Ac)P (B) = P (Ac)P (Bc).

1.41 a.

P( dash sent | dash rec)
 =

Second Edition 1-9

P( dash rec | dash sent)P( dash sent)

P( dash rec | dash sent)P( dash sent) + P( dash rec | dot sent)P( dot sent)

=

(2/3)(4/7)
(2/3)(4/7) + (1/4)(3/7)

= 32/41.

b. By a similar calculation as the one in (a) P (dot sent|dot rec) = 27/434. Then we have

P( dash sent|dot rec) =~~16~~

Given that dot-dot was received, the distribution of the four

43 .

possibilities of what was sent are

Event Probability

dash-dash (16/43)2

dash-dot (16/43)(27/43)

dot-dash (27/43)(16/43)

dot-dot (27/43)2

1.43 a. For Boole’s Inequality,

∑ ∑

P (∪ni=1)≤ P (Ai) − P2 + P3 + · · · ± Pn ≤ P (Ai)

i=1 i=1

since Pi ≥ Pj if i ≤ j and therefore the terms −P2k + P2k+1 ≤ 0 for k = 1, . . . ,~~n−1~~ when

2

n is odd. When n is even the last term to consider is −Pn ≤ 0. For Bonferroni’s Inequality
apply the inclusion-exclusion identity to the Aci, and use the argument leading to (1.2.10).

b. We illustrate the proof that the Pi are increasing by showing that P2 ≥ P3. The other
 arguments are similar. Write

∑ ∑ ∑

P2 = P (Ai ∩ Aj ) =

1≤i<j≤n i=1 j=i+1

P (Ai ∩ Aj )
[ ]

∑ ∑ ∑

= P (Ai ∩ Aj ∩ Ak ) + P (Ai ∩ Aj ∩ (∪k Ak )c)

i=1 j=i+1 k=1

Now to get to P3 we drop terms from this last expression. That is

[ ]

∑ ∑ ∑

P (Ai ∩ Aj ∩ Ak ) + P (Ai ∩ Aj ∩ (∪k Ak )c)

i=1 j=i+1 k=1

[ ]

∑ ∑ ∑

≥ P (Ai ∩ Aj ∩ Ak )

i=1 j=i+1 k=1

∑ ∑ ∑ ∑

≥ P (Ai ∩ Aj ∩ Ak ) = P (Ai ∩ Aj ∩ Ak ) = P3.

i=1 j=i+1 k=j+1 1≤i<j<k≤n

The sequence of bounds is improving because the bounds P1, P1 −P2 +P3, P1 −P2 +P3 −P4 + P5,..., are getting smaller since Pi ≥ Pj if i ≤ j and therefore the terms −P2k + P2k+1 ≤ 0. The lower bounds P1 − P2, P1 − P2 + P3 − P4, P1 − P2 + P3 − P4 + P5 − P6, . . ., are getting bigger since Pi ≥ Pj if i ≤ j and therefore the terms P2k+1 − P2k ≥ 0.

n

n

n−1

n

n−1

n

n

n−1

n

n

n−1

n

n

n−2

n−1

n

1-10 Solutions Manual for Statistical Inference

c. If all of the Ai are equal, all of the probabilities in the inclusion-exclusion identity are the

same. Thus ) )

P1 = nP(A), P2 =

(n

2

(n

P (A), . . . ,Pj = P (A),

j

and the sequence of upper bounds on P (∪iAi) = P (A) becomes

[ ) )]

P1 = nP(A), P1 − P2 + P3 = n−

(n

2

(n

+ P (A), . . .

3

which eventually sum to one, so the last bound is exact. For the lower bounds we get

[

P1 − P2 = n−

)]

(n

2

P (A), P1 − P2 + P3 − P4 =

[ ) ) )]

(n (n (n

n− + − P (A), . . .

2 3 4

which start out negative, then become positive, with the last one equaling P (A) (see Schwa-

ger 1984 for details).

1.44 P (at least 10 correct|guessing) =∑20

(20)(1)k (3)n−k
 = .01386.

k=10 k 4 4

1.45 X is finite. Therefore B is the set of all subsets of X . We must verify each of the three properties

in Definition 1.2.4. (1) If A ∈ B then PX (A) = P (∪xi∈A{sj ∈ S : X(sj) = xi}) ≥ 0 since P

is a probability function. (2) PX (X ) = P (∪mi=1{sj ∈ S : X(sj) = xi}) = P(S) = 1. (3) If

A1,A2,... ∈ B and pairwise disjoint then

⋃

PX(∪∞k=1Ak ) = P( {∪xi∈Ak {sj ∈S:X(sj )=xi}})

k=1

∑ ∑

= P (∪xi∈Ak {sj ∈S:X(sj )=xi}) = PX(Ak),

k=1 k=1

where the second inequality follows from the fact the P is a probability function.

1.46 This is similar to Exercise 1.20. There are 77 equally likely sample points. The possible values of

X3 are 0, 1 and 2. Only the pattern 331 (3 balls in one cell, 3 balls in another cell and 1 ball in a(7)(7)(4)

third cell) yields X3 = 2. The number of sample points with this pattern is

2 3 3

5 = 14,700.

So P (X3 = 2) = 14,700/77 ≈ .0178. There are 4 patterns that yield X3 = 1. The number of

sample points that give each of these patterns is given below.

pattern number of sample points

(7)

34 7 6 = 1,470

(3)(6)(4)(2)

7

322 7 = 22,050

(3) (4)(5)2

7

3211 7 6 2! = 176,400

(3)(6) 2

31111 7

7

3 4

4! = 88,200

288,120

So P (X3 = 1) = 288,120/77 ≈ .3498. The number of sample points that yield X3 = 0 is

77 − 288,120 − 14,700 = 520,723, and P(X3 = 0) = 520,723/77 ≈ .6322.

1.47 All of the functions are continuous, hence right-continuous. Thus we only need to check the
 limit, and that they are nondecreasing

a. limx→−∞1(1 2

d

)
 = 0, limx→∞ 1 2 +π tan−1(x)=2 +π (2)=1,and

dx

2+πtan−1(x))n−11+x2=> 0, so(= F (x) is increasing.

b. See Example 1.5.5.

c. limx→−∞ e−e−x = 0, limx→∞ e−e−x = 1,~~d~~dx e−e−x = e−xe−e−x > 0.

d. limx→−∞(1 − e−x) = 0, limx→∞(1 − e−x) = 1,~~d~~dx (1−e−x)=e−x > 0.

∞

∞

∞

2

2

−π

1−ϵ

Second Edition
1−ϵ

1−ϵ

1−ϵ

1-11

e. limy→−∞ 1+e−y = 0, limy→∞ ϵ + 1+e−y = 1,~~d~~dx (1+e−y ) = ~~(~~~~1−ϵ)e−y~~(1+e−y )2 > 0 and~~d~~dx (ϵ+ 1+e−y ) >

0, FY (y) is continuous except on y = 0 where limy↓0(ϵ + continuous.

1−ϵ

1+e−y )

= F(0). Thus is FY (y) right

1.48 If F (·) is a cdf, F (x) = P (X ≤ x). Hence limx→∞ P (X ≤ x) = 0 and limx→−∞ P (X ≤ x) = 1.
 F (x) is nondecreasing since the set {x : X ≤ x} is nondecreasing in x. Lastly, as x ↓ x0,
 P (X ≤ x) → P (X ≤ x0), so F (·) is right-continuous. (This is merely a consequence of defining
 F (x) with “ ≤ ”.)

1.49 For every t, FX (t) ≤ FY (t). Thus we have

P (X > t) = 1 − P (X ≤ t) = 1 − FX (t) ≥ 1 − FY (t) = 1 − P (Y ≤ t) = P (Y > t). And for some t∗, FX (t∗) < FY (t∗). Then we have that

P (X > t∗) = 1 − P (X ≤ t∗) = 1 − FX (t∗) > 1 − FY (t∗) = 1 − P (Y ≤ t∗) = P (Y > t∗).

1.50 Proof by induction. For n = 2

∑

tk−1 = 1 + t =

k=1

1−t2

1−t

Assume true for n, this is∑nk=1 tk−1 =

∑ ∑

~~1−t~~ .Thenforn+1

1−tn 1−tn+tn(1−t) 1−tn+1

tk−1 = tk−1 + tn =

k=1 k=1

1−t

+tn = = ,

1−t 1−t

where the second inequality follows from the induction hypothesis.

1.51 This kind of random variable is called hypergeometric in Chapter 3. The probabilities are

obtained by counting arguments, as follows.

x

0

fX(x) = P(X = x)

(5)(25)∕(30)

≈ .4616

0 4 4

(5)(25)∕(3 )

1

0

1 3 4

≈ .4196

(5)(25)∕(3 )

2

0

2 2 4

≈ .1095

(5)(25)∕(3 )

3

0

3 1 4

≈ .0091

(5)(25)∕(3 )

4

0

4 0 4

≈ .0002

The cdf is a step function with jumps at x = 0, 1, 2, 3 and 4.

1.52 The function g(·) is clearly positive. Also,

∫ ∞ ∫ ∞

g(x)dx =

x0 x0

f (x) −F (x0)
1−F (x0)dx=1−F (x0)

= 1.

1.53 a. limy→−∞ FY (y) = limy→−∞ 0 = 0 and limy→∞ FY (y) = limy→∞ 1 −~~1~~ = 1. For y ≤ 1,

y2

FY (y) = 0 is constant. For y > 1,~~d~~dy FY (y)=2/y3 >

FY is nondecreasing. Therefore FY is a cdf.
 {

2/y3 if y > 1

0, so FY is increasing. Thus for all y,

b. The pdf is fY (y) =~~d~~dy FY (y)=

0 if y ≤ 1.

c. FZ (z) = P (Z ≤ z) = P (10(Y − 1) ≤ z) = P (Y ≤ (z/10) + 1) = FY ((z/10) + 1). Thus,

FZ(z) =

{0 (

1−

)

1

[(z/10)+1]2

if z ≤ 0

if z > 0.

2

1−tn

n+1

n

1-12

1.54 a.

∫π/2

0

∫ ∞

Solutions Manual for Statistical Inference

sin xdx = 1. Thus, c = 1/1 = 1.
 ∫0

b. −∞ e−|x|dx = −∞ exdx+ 0

1.55

e−xdx = 1 + 1 = 2. Thus, c = 1/2.

∫3

P (V ≤ 5) = P (T < 3) =

0

For v ≥ 6,
 ( )

1

1.5e−t/1.5 dt=1−e−2.

∫ v

2

Therefore,

P (V ≤ v) = P (2T ≤ v) = P T ≤

{0

v

2

1

=

0 1.5e−t/1.5 dt=1−e−v/3.

−∞ < v < 0,

P (V ≤ v) = 1−e−2 0≤v<6,

1−e−v/3 6≤v