Some exercises & examples from Ch 2

Ex 2.2.17
$P(A)=.3$, $P(B)=.5$ is given

a) There are requests that do not relate to neither SPSS nor SAS (maybe R!)
b) $P(A') = 1-P(A) = 1- .3 = .7$
c) First recall $\cup$ means "union" (inclusive or), $\cap$ means "intersection" (and).
$P(A\cup B)= P(A) + P(B) – P(A\cap B)$ but in the context of the question it is reasonable to assume that $P(A\cap B)=0$, so $P(A\cup B)=.3+.5=.8$
d) $P(A' \cap B') = ?$ First look at the Venn diagram – if a point of the sample space is outside A AND outside B, it must be outside $A\cup B$, and the converse is also true.
So $A' \cap B'$ is the same set as $(A\cup B)'$. Thus $P(A' \cap B')=1-P(A\cup B)=1-.8=.2$

Ex 2.2.24

It is always true that $B=(B\cap A) \cup(B\cap A')$ (Think Venn to see this).
If A is contained in B, then $B\cap A = A$, so $B=A\cup(B\cap A')$ as in the hint.
But A and $(B\cap A')$ are disjoint, so $P(B)=P(A\cup(B\cap A'))=P(A) + P(B\cap A') \geq P(A)$ qed
For the next part of the question ...
Note that $A\cap B$ is contained in A which is contained in $A\cup B$
So $P(A\cap B) \leq P(A) \leq P(A\cup B)$

Note: The relationship $B=(B\cap A) \cup(B\cap A')$ is often useful for calculating P(B), since often events like $B\cap A$ or $B\cap A'$ are simpler than B.

Example 2.23

We have 25 printers: 10 laser and 15 inkjet. Select 6 at random. What is the chance of 3 laser and 3 inkjet in this selection? Each possible selection of 6 printers is equally likely, so to get the probability, we need to count the number of select-6 instances that result in 3+3, and divide this by the number of ways of selecting 6. (Why? Because each possible random selection is mutually exclusive of each other one, and so the probabilities of the union of all selections resulting in 3+3 is $1/n + 1/n + ...+1/n$ where n is the number of possible random samples, and the number of terms is the number of samples resulting in 3+3. The n here is the number of ways of selecting 6. So the sum of the series is the same as the number of 3+3 samples divided by n.)

How many ways can we get 3 laser printers from the 10? $C_{3,10}$ (we compute shortly)
How many ways can we get 3 inkjet printers from the 15? $C_{3,15}$ (we compute shortly)
How many ways to get both? Use the mn rule: $C_{3,10}$ times $C_{3,15}$
How may ways to get a sample of 6 from the 25? $C_{6,25}$

So the required probability is ($C_{3,10}$ times $C_{3,15}$)/ $C_{6,25}$

Evaluate this by hand as on p 73, or use R

Unfortunately, you need to know that the calculation is is a special case of the hypergeometric distribution – one of the dozen or so probability models we will study. Then, with this knowledge if you give the command

help.search("hypergeometric") you will get this:

Hypergeometric {stats}

The Hypergeometric Distribution

Description

Density, distribution function, quantile function and random generation for the hypergeometric distribution.

Usage

```
dhyper(x, m, n, k, log = FALSE)
phyper(q, m, n, k, lower.tail = TRUE, log.p = FALSE)
qhyper(p, m, n, k, lower.tail = TRUE, log.p = FALSE)
rhyper(nn, m, n, k)
```

Arguments

`x, q`	vector of quantiles representing the number of white balls drawn wit balls.
`m`	the number of white balls in the urn.
`n`	the number of black balls in the urn.
`k`	the number of balls drawn from the urn.
`p`	probability, it must be between 0 and 1.
`nn`	number of observations. If `length(nn) > 1`, the length is taken to b
`log, log.p`	logical; if TRUE, probabilities p are given as log(p).
`lower.tail`	logical; if TRUE (default), probabilities are *P[X <= x]*, otherwise, *P*

Details

The hypergeometric distribution is used for sampling *without* replacement. The density of this distribution with parameters `m`, `n` and `k` (named *Np*, *N-Np*, and *n*, respectively in the reference below) is given by

$$p(x) = choose(m, x)\ choose(n, k\text{-}x) / choose(m+n, k)$$

for $x = 0, ..., k$.

Value

`dhyper` gives the density, `phyper` gives the distribution function, `qhyper` gives the quantile function, and `rhyper` generates random deviates.

References

Johnson, N. L., Kotz, S., and Kemp, A. W. (1992) *Univariate Discrete Distributions*, Second Edition. New York: Wiley.

Examples

```
m <- 10; n <- 7; k <- 8
x <- 0:(k+1)
rbind(phyper(x, m, n, k), dhyper(x, m, n, k))
all(phyper(x, m, n, k) == cumsum(dhyper(x, m, n, k)))#
FALSE
## but error is very small:
signif(phyper(x, m, n, k) - cumsum(dhyper(x, m, n, k)),
dig=3)
---------------------------------------
```

So what we need is dhyper which gives probabilities for this distribution:

```
> dhyper(3,15,10,6)
[1] 0.3083004
```

and for the rest of the question

```
> dhyper(3,15,10,6)
[1] 0.3083004
> dhyper(4,15,10,6)
[1] 0.3468379
> dhyper(5,15,10,6)
[1] 0.1695652
> dhyper(6,15,10,6)
[1] 0.02826087
> result=.3083+.3468+.1696+.0283
```

```
> result
[1] 0.8530
```

The hypergeometric distribution arises whenever a population of items of two kinds is selected at random, without replacement. More on pp 129 ff (in ch 3).

Ex 2.3.31

9 B
27M

How many ways to play a B and an M, in that order. Use mn rule. 9*27 = 243 ways (or think, each B has 27 followers, and there are 9 Bs).

next part:

9B
27M
15S

9*27*15/365 = 9.99 or approx 10 years without repetition.

Ex 2.3.35

There are only two sequences where A is always ahead: AAABB and AABAB.
All orders of the 3 As and 2 Bs are equally likely and there are a total of ??? orders.
The required probability is 2/???. We need the "???"

The answer is ???=5!/(3! 2!) = 10

The important theorem underlying this can be explained in two ways:

i) the letters A,B,C,D,E can be ordered in 5! ways. But if we identify A, with D and E and identify B with C, we lose the distinction between 3! * 2! orders, so there are 5!/(3! 2!) = 10

ii) The orders of the 3 As (or the 2 Bs) can be specified by selecting the order number of each one. So for example if the As had orders 1,2,4, the result would be AABAB.
But the number of ways of doing this is $C_{3,5}$ = 10. (We could also use $C_{2,5}$=10).

This problem is a special case of another famous theorem called the Ballot Theorem:
In general if there are m As and n Bs, the probability that A will always be ahead of B is (m-n)/(m+n)
If you are astonished at the simplicity of this result, you will realize why this is a famous theorem!

Ex 2.4.43

Five card poker, aces high or low, straight is 5 denominations in a row.
P(straight) = ?
It seems to matter if the first card chosen is in the range 5 to 10 or not, for then the straight can develop either way. So lets first find the probability of a 10-high straight. That is 6,7,8,9,10. Well, there are 4 of each of 13 denominations. We need 1 of the 4 6s, 1 of the 4 7s, etc and 0 of the 36 others.
We could write
$((C_{1,4})$^5 times $C_{0,36})/C_{5,52}$ Now $C_{0,36} = 1$.

In R this looks like

```
> (4^5)/choose(52,5)
[1] 0.0003940038
```

so the probability of a 10-high straight is .000394.

The next question asks to find the probability of any straight.

But if the prob of 6,7,8,9,10 is .000394, so is the probability of
A,2,3,4,5 or 2,3,4,5,6, ..., or 10,J,Q,K,A.

But there are 10 such straights. So the probability of one of them occurring is 10*.000394 =.00394

A straight flush is a special straight that has all suits the same. Consider the one straight 10,J,Q,K,A. We know there are 4^5 straights like this, and 4 of them are straight flushes. So of all the straights, a proportion 4/(4^5) are flushes. The probability of a straight flush must be .00394 times 4/(4^5) = .00001539

(KLW 070117)