

Mark-recapture Distance Sampling

Workshop, 24-25 August 2020

Centre for Research into Ecological and Environmental Modelling

Hidden Markov Line Transect Model Exercise

1 Objectives

The aim of this exercise is to illustrate the process of drawing inference using hidden Markov line transect models (HMLTMs):

1. Understand the data required for inference,
2. Explore a number of detection models,
3. Illustrate model selection,
4. Illustrate how confidence intervals are obtained,
5. Compare with conventional line transect methods using “correction factors”.

2 The Data

The data used in this exercise were gathered on a shipboard survey of beaked whales in the Alboran sea in 2008, from a vessel moving at about 3.5 m/s. Beaked whales have long dives and spend a lot of time underwater and unobservable. Focal follows of whales gave estimates of mean times available (at surface) and unavailable (diving) in a single dive cycle of 122 seconds and 1,580 seconds, respectively, with standard errors of these means of 9.62 seconds and 134.9 seconds, respectively.

A total of 81 groups were detected. The sightings data contain forward distance (y) as well as perpendicular distance (x), and also group size ($size$), Beaufort sea state (bf) and the height of the observer who made the detection (ht). We are going to focus on x and y only here. The survey was not a MRDS survey and so the data contains no duplicate data.

```
library(hmltm) # load the library
data("beaked.ship") # load the data (contained in the library)
knitr::kable(head(beaked.ship))
```

stratum	area	transect	L	x	y	size	bf	object	ht
1	1000	1	100	88	351	3	3	1	3.60
1	1000	1	100	798	1228	7	7	2	17.20
1	1000	1	100	0	1212	2	2	3	17.20
1	1000	1	100	327	146	2	2	4	4.75
1	1000	1	100	2188	1263	4	4	5	17.20

stratum	area	transect	L	x	y	size	bf	object	ht
1	1000	1	100	75	279	5	5	6	4.75

3 Conventional Distance Sampling Analysis

Let's start by fitting a conventional distance sampling model to the data. To do that we first have to rename some columns in the data frame `beaked.ship` to comply with the conventions of the `Distance` package. We will put the data frame with the new column names in something called `bkdsdat`:

```
bkdsdat = beaked.ship
names(bkdsdat)[1:4] = c("Region.Label", "Area", "Sample.Label", "Effort")
names(bkdsdat)[which(names(bkdsdat)=="x")] = "distance"

knitr::kable(head(bkdsdat))
```

Region.Label	Area	Sample.Label	Effort	distance	y	size	bf	object	ht
1	1000	1	100	88	351	3	3	1	3.60
1	1000	1	100	798	1228	7	7	2	17.20
1	1000	1	100	0	1212	2	2	3	17.20
1	1000	1	100	327	146	2	2	4	4.75
1	1000	1	100	2188	1263	4	4	5	17.20
1	1000	1	100	75	279	5	5	6	4.75

To fit a CDS model, we need the `Distance` package and we need to specify the units of the distances, transect lengths and areas, which are as follows:

```
library(Distance)
conversion.factor <- convert_units("meter", "kilometer", "Square kilometer")
```

Lets try a half-normal and a hazard-rate model, and compare their AICs:

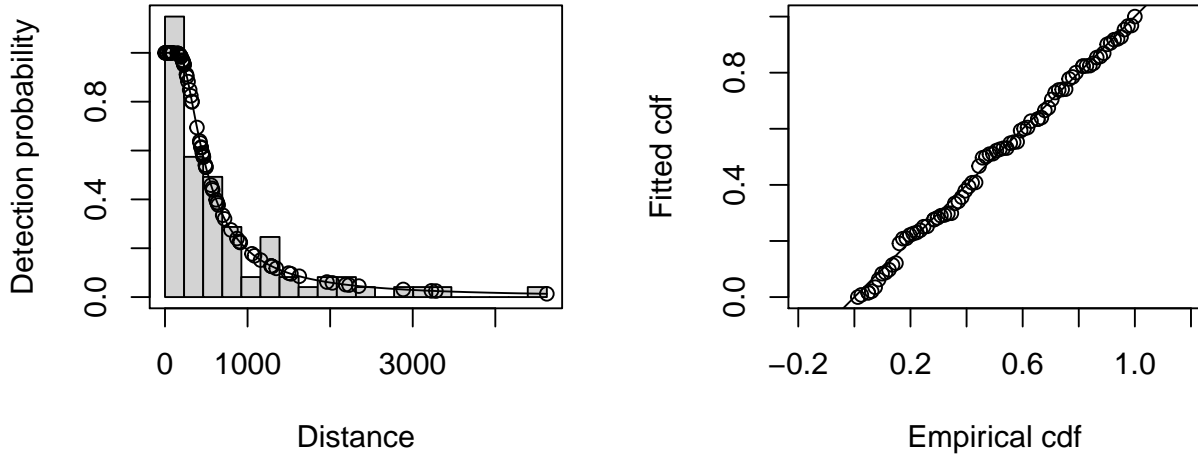
```
bk.hn <- ds(data=bkdsdat, key="hn", adjustment=NULL, convert.units=conversion.factor)
bk.hr <- ds(data=bkdsdat, key="hr", adjustment=NULL, convert.units=conversion.factor)
knitr::kable(AIC(bk.hn,bk.hr))
```

	df	AIC
bk.hn	1	1259.314
bk.hr	2	1232.746

The hazard-rate model is clearly preferable, so lets check that it seems adequate before proceeding:

```
cutpoints = seq(0,max(na.omit(bkdsdat$distance)),length=21)
par(mfrow=c(1,2))
plot(bk.hr, breaks=cutpoints, main="Half normal model, beaked shipboard survey")
gof_ds(bk.hr)
```

Half normal model, beaked shipboard survey



```
##
## Goodness of fit results for ddf object
##
## Distance sampling Cramer-von Mises test (unweighted)
## Test statistic = 0.0291797 p-value = 0.978784
```

That looks OK, so we will proceed using this model. We know that $g(0)$ is less than 1 (because of how long the whales dive), so the CDS estimates will be negatively biased, but lets look at the total abundance estimate anyway:

```
cds.N = c(est=bk.hr$dht$individuals$N$Estimate[3],
          lcl=bk.hr$dht$individuals$N$lcl[3],
          ucl=bk.hr$dht$individuals$N$ucl[3])
caption="CDS Beaked whale abundance estimates"
knitr::kable(t(round(as.data.frame(cds.N))),caption=caption)
```

Table 4: CDS Beaked whale abundance estimates

	est	lcl	ucl
cds.N	361	232	560

4 CDS with Correction Factors

Lets now calculate various correction factors and use them to “correct” for $g(0) < 1$. We will do this for the simple (instantaneous) correction factor (whihc is just the proportion of time available), McLaren’s factor, and Laake’s factor. The `hmltm` package has functions that allow you to do this, given the mean times available and unavailable, and (optionally) estimates of the CVs of these times. McLaren’s and Laake’s factors require us to specify a forward distance at which animals are first detectable if they are available (we will call this y_{\max}).

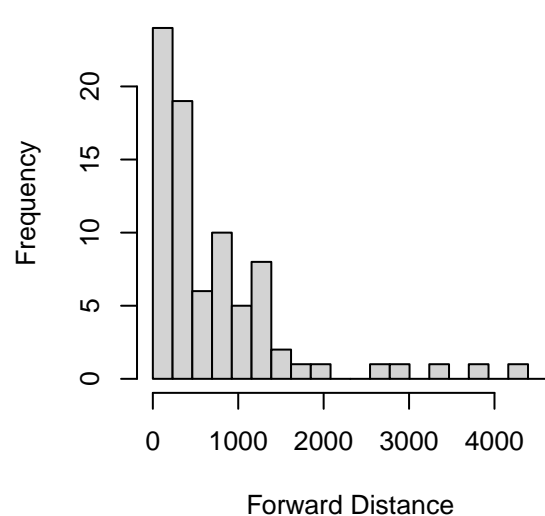
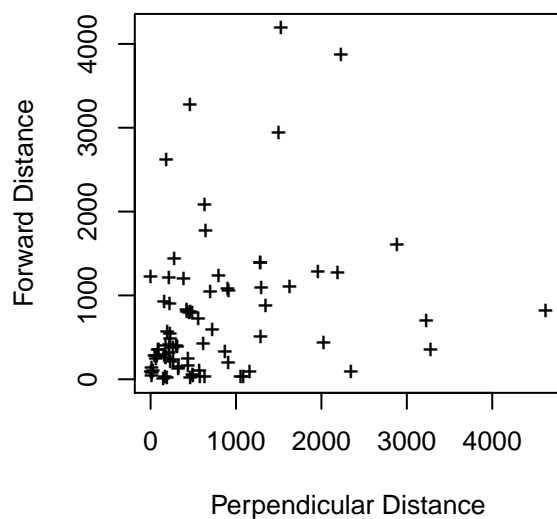
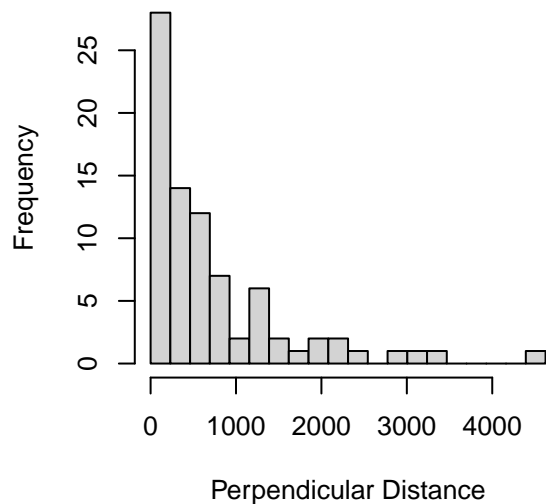
Lets plot the data in x - and y - dimensions to figure out what a reasonable y_{\max} would be

```
xcuts = seq(0,max(na.omit(beaked.ship$x)),length=21)
ycuts = seq(0,max(na.omit(beaked.ship$y)),length=21)
```

```

layout(matrix(1:4,nrow=2))
hist(beaked.ship$x,breaks=xcuts,xlab="Perpendicular Distance",main="")
plot(beaked.ship$x, beaked.ship$y, pch="+",xlab="Perpendicular Distance",
      ylab="Forward Distance")
plot(0,0,xaxt="n",yaxt="n",bty="n",xlab="",ylab="",type="n")
hist(beaked.ship$y,breaks=xcuts,xlab="Forward Distance",main="")

```



From this, a `ymax` equal to 4,500 seems reasonable, so we will go with that. The `hmltm` package has a function `make.hmm.pars.from.Et` that constructs the HMM parameter object needed for inference from the mean times unavailable, mean time available, `ymax`, and survey platform speed:

```

ymax = 4500 # max forward distance can detect animal

Eu=1580 # mean time UNavailable -per dive cycle
Ea=122 # mean time available per dive cycle
spd = 3.5 # ship speed in m/s
# Calculate time in view, in minutes

```

```
tiv = ymax/spd / 60 # minutes in view

# Make availability parameters object
availpars = make.hmm.pars.from.Et(Eu=Eu, Ea=Ea, seEu=134.9, seEa=9.62)
```

From this we see that the time an animal is within detectable distance is 21.43 minutes, while the mean time UNavailable is 26.33 minutes and the mean time available is 2.03 minutes. Clearly $g(0)$ is going to be a lot less than 1 - but how much less? Let's calculate the naive estimate and ask McLaren and Laake. Package `hmltm` provides the functions `simple.a`, `mclaren.a`, and `laake.a` to do the calculations, using objects like `availpars`, which we created above, as input:

```
simple.cf = simple.a(availpars)
laake.cf = laake.a(availpars,spd=spd, ymax=ymax)
mclaren.cf = mclaren.a(availpars,spd=spd, ymax=ymax)
# put them in a list for convenience
cf = data.frame(simple=simple.cf$mean,
                 mclaren=mclaren.cf$mean,
                 laake=laake.cf$mean)
knitr::kable(round(cf,4),caption="Availability correction factors")
```

Table 5: Availability correction factors

simple	mclaren	laake
0.0717	0.8271	0.5886

Lets use these to “correct” the CDS abundance estimates:

```
# Point estimates of total abundance and 95% CI with each correction method
simple.N = cds.N/cf[, "simple"]
mclaren.N = cds.N/cf[, "mclaren"]
laake.N = cds.N/cf[, "laake"]
Nests = t(data.frame(simple=round(simple.N),mclaren=round(mclaren.N),laake=round(laake.N)))
knitr::kable(Nests, caption="Corrected abundance estimates")
```

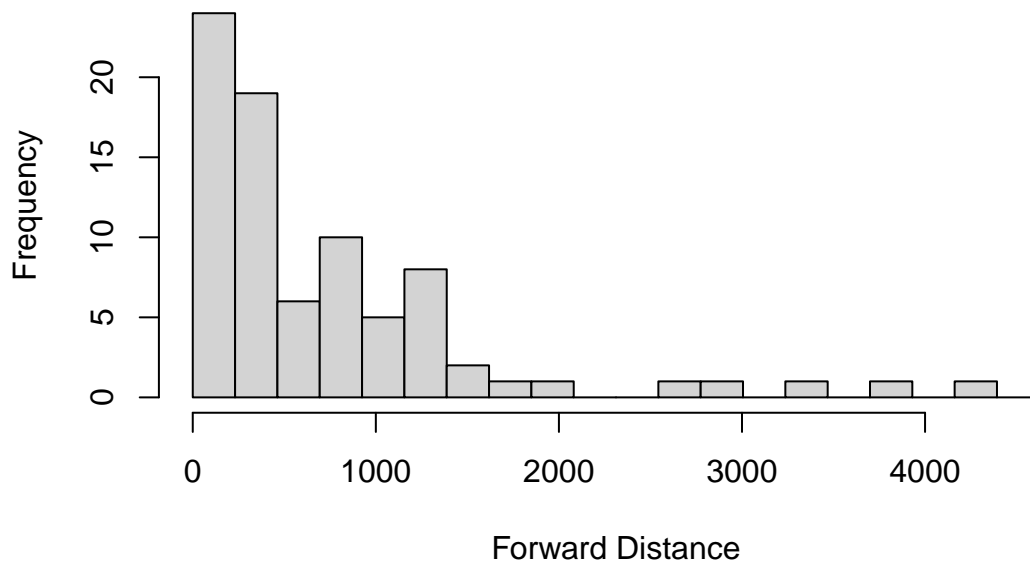
Table 6: Corrected abundance estimates

	est	lcl	ucl
simple	5031	3243	7806
mclaren	436	281	676
laake	613	395	951

There are big differences between the methods! Which one to believe?

Lets take another look at the forward distance data and think about what this implies for the validity of each of the above correction methods:

```
hist(beaked.ship$y,breaks=xcuts,xlab="Forward Distance",main="")
```



4.1 The simple correction factor

The simple method estimates the probability that a whale is available at a randomly chosen *instant*. But here the whales are available (to some extent) for 21.43 minutes, which is very far from an instant, and so this simple method estimate is going to be way too small, and the corresponding “corrected” abundance estimate way too big.

4.2 McLaren’s correction factor

McLaren’s method estimates the probability that a whale is available **at least once** while within detectable distance, but it does this in a way that can allow the probability to go above 1, and Laake’s method fixes this problem, so Laake’s method is better than McLaren’s and we should prefer it to McLaren’s.

4.3 Laake’s correction factor

Laake’s method estimates the probability that a whale is available **at least once** while within detectable distance, and does so in a sensible way, but is this the appropriate correction factor?

Look at the forward distance plot and consider the fact that *where* the animal becomes available between forward distance 0 and 4500 m. If it becomes available anywhere farther ahead than about 1,500 m, there is only a very small chance that it will be detected (we know this since very few detections occurred beyond about 1,500 m). So you would expect that an appropriate correction factor should be smaller than the probability that a whale is available **at least once**, and hence that Laake’s correction factor is too big, and the associated abundance estimate too small.