Cormack-Jolly-Seber Models – notes

WIS 4601 - Teaching Demo

Last compiled: 2025-04-17

## Title slide

Today we are going to learn about one of the most famous models in wildlife biology, the Cormack-Jolly-Seber model, which uses capture-mark-recapture data to estimate survival of animals would accounting for a pesky nuisance variable, imperfect detection.

I’m going to walk you through some general concepts related to why we need this model, study design and assumptions, and how we can conceptualize and build this model this model hierarchically using probability functions. Then I will show examples of how to perform these analyses using both frequentist and Bayesian methods in R. As always, we will use simulated data in class to make sure the model is estimating parameters we know to be ‘truth’, and then in lab later this week, you all will work through analysis with both simulated and real data.

## Review

A quick review before we dive into today’s topic. Remember that for the final few weeks of the semester, our goal is to work toward building an ‘integrated population model’. Over the last two weeks, we have learned how to (1) estimate population size using count data and hierarchical state-space models, the model in this central grey box, but then we also learned (2) how to estimate population productivity using data describing reproduction using Poisson regression models, and then (3) last class we learned how to estimate survival using ‘known-fate data’ with a Kaplain-Meier model. We are learning how to build each of these individual models for abundance, reproduction, and survival, so we can put them all together into integrated population models at the end of the semester.

But radiotags or GPS devices are expensive, and we don’t have ‘known fate’ data.

## Review 2

Instead, we rely on capture-mark-recapture data to estimate survival. These mark-recapture data, M, inheritly are challenged by this pesky nuisance parameter, imperfect detection, annotated P in this graph. today we will learn how to build CJS models that can estimate survival while accounting for imperfect detection.

## Readings

I asked you all to read Chapter 7 of Kery & Schaub 2012 before class, which provides a simple introduction to these models and how to fit them. The other two chapters also provide context on implementation in both Bayesian and frequentist methods - providing these here mostly for reference.

For today’s class, you can access the slides, code, and data on the course website, as always. Navigate to <https://brianfolt.github.io/NRES_710/cjs_models.html>, and then there are links on that page to access all the material.

## Why do we need CJS models?

We talked about this earlier this semester when we learned about closed-population models to estimate population size, the Lincoln-Peterson model and state-space abundance models…. but most animals are difficult to observe or capture.

Here’s a picture of a Great Basin landscape, that has an animal in it… it’s not exactly a stumper of a picture….

## Why do we need CJS models? 2

The animal is front and center in the picture, but it’s color and shape blends in with the sand, and it’s also hidden in the shade of a grass clump….

## Why do we need CJS models? 3

It’s a Desert Horned Lizard! The species relies almost entirely on crypsis to avoid getting eaten by avian predators.

I saw this one hiding there, but I’m almost certain there were other individuals out there that day that I didn’t see…

The bottom line is: **a hallmark problem of wildlife ecology is that when we survey for animals, they are never detected perfectly**. If we want to get unbiased and precise estimates of population size or survival, we need to account for imperfect detection. If we do not account for imperfect detection, we will often:

* underestimate population size
* and underestimate survival

## Study design

As we have seen with closed population models, we can use capture-mark-recapture data to account for imperfect detection when estimating population size.

Basic study design involves going out, capturing animals, uniquely marking them with some way that will provide a permanent form of identification of that individual into the future, and then let the animal go alive!

There are three examples of different marking ‘technologies’ in these pictures: alligator with a tail tag, an alligator snapping turtle with unique combinations of screws being drilled into the marginal scutes of the carapace, or a tag being banded onto the leg of a large bird.

To estimate survival though, we need to repeat these sampling occasions through time. We might survey the study area every year, and then we can use these data to estimate *annual survival probability* through time.

## Study design 2

There are a few considerations when designing these studies:

* Fixed study area: area should not change through time. Fairly simple & standard approach. If the study area increases, site fidelity will increase, and apparent survival will increase.
* Study area is ~large. We want it to encompass a relatively large sample of individuals, and preferably those individuals will not leave the study area.
* Samples are ~instantaneous in time. We generally don’t want to take our sample at different times of the year, or else we might bias our estimation of survival probabilities.
* Individuals are a random sampled of the population during each occasion

I will revisit these design features later on.

## Open population models 1

We previously discussed ‘**closed population models**’, where we assumed the population was closed to any change in $\large N$ during our study

* There are no births, deaths, emigration, or immigration during the study.
* Any $\large 0$ in the capture histories was due to detection error – the animal was present in population, but it simply went undetected. It was a lizard that was hiding in the grass, and we missed it.

## Open population models 2

Open population models let us relax the assumption of population closure.

Between sampling occasions, individuals can enter the population through birth or immigration, or leave the population by dying or emigrating away.

Zeros in our data can result from animals (1) being present and undetected, OR (2) being not present because they either died or emigrated away.

Different open population models can be used to estimate **survival**, **recruitment**, and **movement** – but our focus today is on survival estimation.

## Cormack-Jolly-Seber (CJS) models

CJS models are open-population **state-space models** to estimate **survival**.

We previously talked about **state-space models** for population size… these were models that estimate a true, underlying, “latent” state of populations (THE STATE PROCESS), while accounting for imperfect detection (OBSERVATION PROCESS). We can do something similar with CJS models.

For CJS survival models, we are interested in estimating the true underlying, often-unobservable state of individuals: each circle represents a time step, and whether individuals are alive (blue) or dead (white). Each individual has a stochastic probability of surviving and staying alive – but if they don’t survive, then they die, and they stay dead into the future – this is deterministic.

## Cormack-Jolly-Seber (CJS) models 2

Overlaying this true, latent, or ‘hidden’ state, is the observation process. Again, when we go out into nature, animals are not always detected, even if they are alive and present. The red circles indicate situations where an animal was alive, and we observed it. But sometimes we do not. This is a stochastic process that can vary randomly. However, once an animal is dead, this becomes deterministic – we cannot observe a dead animal because it’s ‘dead and gone.’

Using our capture histories through time and a state-space model, we are going to try to estimate these two different, stochastic process: the probability of surviving or dying (the black dashed arrows), and the probability of an alive animal being recaptured (red dashed arrows).

An important feature to mention is that this model is ‘conditioned on the first capture’. When the first capture occurs, we know an animal was certainly alive and observed, so these processes are deterministic. Into the future, the animal may survive and be determined, but the model considers those processes randomly-varying or stochastic.

## Cormack-Jolly-Seber (CJS) models

Data are most frequently organized as ‘capture histories’, which are a series of $\large 1$s and $\large 0$s denoted when individuals are captured ($\large 1$) or not ($\large 0$). The capture histories have dimensions of n x t: each row is for an individual, *n*, and each column is a sampling occasion. So for example, if we sample the population every summer, each column will be our annual sampling occasion, and each entry is whether or not we observed each individual.

Here is an example: here are capture histories for 10 individuals over a 7 year period. The first few animals were recaptured quite a few times, but note individual 7 – it was captured and marked, went undetected in the second year, and then was recaptured again in year 3.

## Cormack-Jolly-Seber (CJS) models

After an animal is marked and released, the next time we go out to sample the population, there are 4 possible scenarios for that animal:

1. Individual **survives** and is **re-captured** (capture history = 11)
2. Individual **survives** but is not **recaptured** (capture history = 10)
3. Individual **dies** and is **not available** for recapture (capture history = 10)
4. Individual **survives** but **leaves the study area** (emigrates) and is **not available** for recapture (capture history = 10)

## Cormack-Jolly-Seber (CJS) models

Without additional data, we can’t distinguish between scenario 3 (death) & scenario (emigration). So, what this model **can** estimate is:

Apparent survival probability – denoted with the greek letter Phi – this is the probability that an individual **survives** and also **stays within the study area**.

We can break this down a bit. Last week we learned how to estimate true survival probability, **S**, using known-fate models. Phi is the true survival times 1 minus the emigration probability, epsilon.

## Cormack-Jolly-Seber models

However, the data we are collecting don’t have any information about true survival or emigration, so we cannot disentangle true survival from emigration… and instead are stuck with estimating apparent survival probability.

The CJS model estimates 2 parameters: - phi – apparent survival probability - p – recapture probability

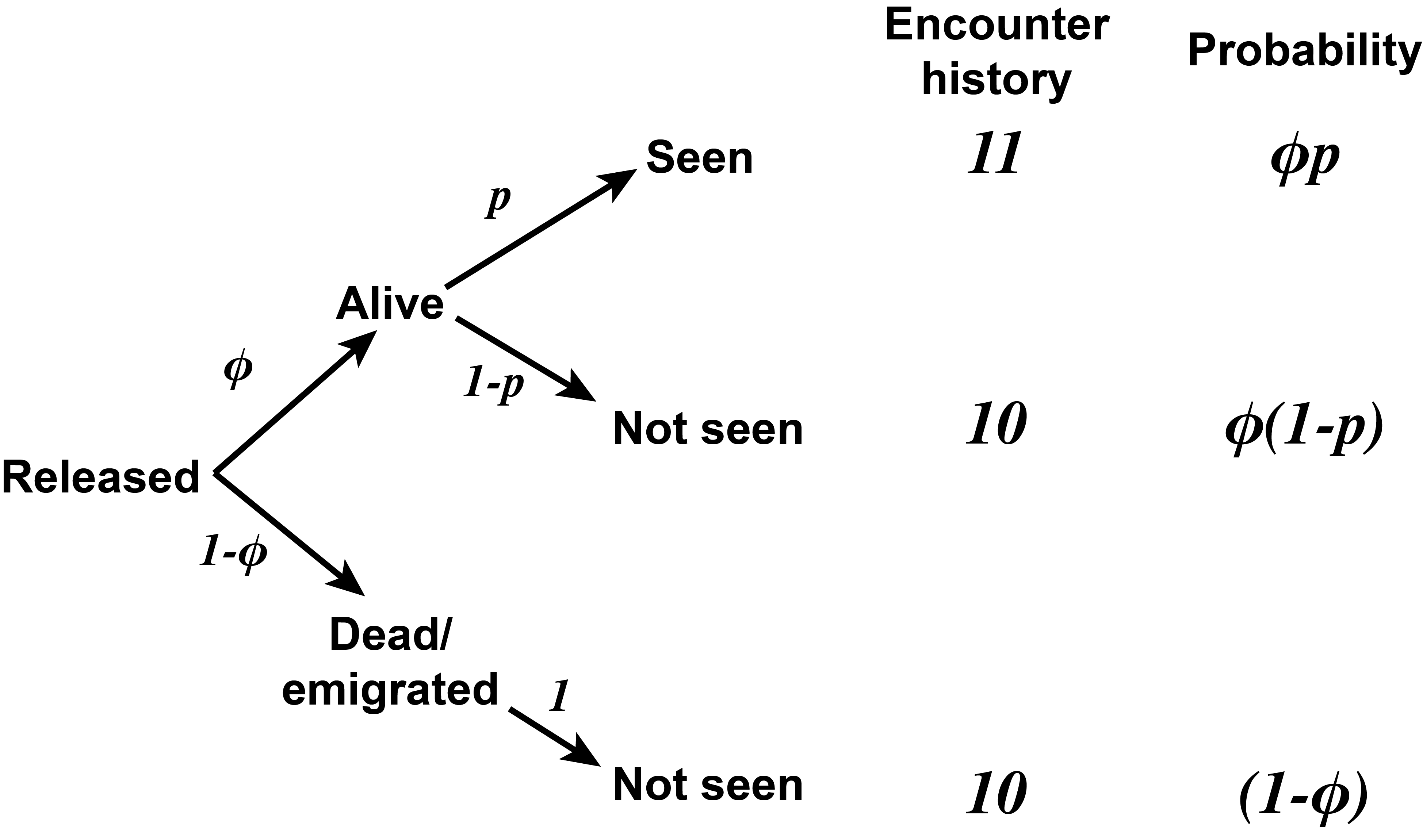
I’ve removed the t subscripts for simplicity, which implies that these probabilities are constant across all sampling occasions.

But, this model can be adapted to having time-varying survival and recapture probabilities, as we’ve done with other models earlier in the semester (i.e., time effects).

## How It Works 1

*Brian write this on the board*

This flow chart is useful to understand how we can use capture histories to disentangle apparent survival, recapture probability, from death and emigration.

Occasion 1 Occasion 2 Encounter History Probability of Data True State True State Observation  


Start on the left hand side. We have captured an animal, marked it, and released it on sampling occasion 1. Over the course of the next year, the animal can either survive and stay in the study area (with probability *phi*), or die or emigrate from the study area (with probability 1 minus *phi*).

If the animal is alive and present, then when we sample the population again, we may recapture it (with recapture probability, *p*), and it is seen again. This animal would then have a capture history of 11 over the first two occasions, and this 11 capture history has a probability of phi \* p.

If this animal is alive but not recaptured, then the capture history is 10, and the probability is phi \* (1-p).

If the animal either dies or emigrates away between the occasions, then we have no chance of recapturing it – it is gone from the population. This also yields an encounter history of 10, but has a different probability: 1 minus phi.

## How It Works - Ind 1

Let’s use this to consider a few different capture-history scenarios. Let’s say we observed an individual over three consecutive occasions: that means it survived two intervals, and was recaptured both times. So we observed a “11” for this animal twice, and using the probabilities from the last slide, the overall probability of this outcome is phi1 \* p2 \* phi2 \* p3.

## How It Works - Ind 2

For individual two, we didn’t observe it on the second occasion, but then recaptured it on the third occasion. This ‘internal zero’ here that is surrounded by 1s has information; since it was bracketed by a 1 on both sides, we know that it survived and stayed in the study area, but went undetected. So the probability of the 10 in this instance is phi(1-p) and then the final 1 is phi\*p – it survived and was detected.

## How It Works - Ind 3

Individual 3 is a little trickier. This animal survived interval 1, and was recaptured on occasion 2 (phi1 \* p2). But we don’t know what happened to it during interval 2: it could have survived interval 2 and gone uncaptured on occasion 3 (1 - p3), OR, died during interval 2 with probability 1 - phi2. Since we don’t know whether it survived or died during interval 2, we sum both of those probabilities together in the second term.

## How It Works - Ind 4

Last, individual 4 is the most confusing. It was captured, marked, and released, and never seen again. It could have:

* died during interval 1 (1 minus phi1), OR
* survived interval 1, gone uncaptured during occasion 2, and died during interval 2, OR
* survived interval 1, gone uncaptured during occasion 2, AND survived interval 2 but went uncaptured during occasion 3.

We can the probabilities of these three outcomes together with this equation, and then simplify it with factoring to the more simple term expressed in the table.

**Takehome message**: This math is confusing, and I don’t want to get too deep into it because it frankly gets more confusing, quickly. But, the takehome message is that:

* different outcomes in our data can be related to distinct probabilities.
* Since these different capture histories have different probabilities, we can use these data to estimate what these detection and survival parameters are using either maximum-likelihood or Bayesian methods.

**Q:** Any questions?

## CJS model as a state-space model

Using the tools we’ve learned this semester, we can write the **CJS model** as a Bayesian **state-space model**. I prefer to teach and implement CJS models with Bayesian approaches, because they are more simple and intuitive. Compared to those complicated likelihood functions on the last slide, I am always impressed by the small amount of JAGS code needed to provide inference for capture-recapture models…

### Process model

In the state-space model, we first have the process model, which is defined by two equations: first, we know animals are alive the first time they are captured, and we denote this with array Z. Z has dimensions i individuals and t years, but each individual’s values for Z start with the first occasion that is it capture: denoted with the vector *f1*.

For capture occasion 1 for each individual, we can only specify it’s initial state. We cannot estimate survival or detection.

Then, for each year after the first capture, an individuals true state, Z\_t, is a Bernoulli draw of it’s previous state, z\_t-1 times phi. So, we if we have lots of data of individuals either being alive or dead through time, we can fit these data with a Bernoulli distribution and estimate the apparent survival probability, phi. Phi is basically a weighted coinflip: survive and stay with probability phi, or die/emigrate with probability 1-phi.

### Observation model

However, we have to account for imperfect detection when we estimate whether or not we think each individual is alive! If we fail to account for imperfect detection, we will underestimate survival. We can do this using the Observation Model. And this is where our data come in!

Our data are y – y is our capture histories, for each individual i and year t. We model whether or not we saw an animal in any given year as a Bernoulli draw of whether it is alive that year, z\_it time recapture probability.

Our data provide information for the samplers to estimate what the most likely probabilities phi and p.

## CJS model with time-variation

As for other models we’ve seen this semester, we can add linear, individual-level covariates or temporal variation to the CJS model using the ‘logit’ link function.

Remember, our parameters of interest here are *probabilities*, which are constrained between 0 and 1. We cannot fit simple linear models to probabilities, because linear models assume that response is continuous and unbounded. Instead, we have to use the ‘logit’ link function – 1 / (1 + exp(-(linear model))) – which unconstrains our betas to be values from negative infinite to positive infinite.

The first equation shows the logit link function, which JAGS does for us in short-hand using the ‘logit()’ command (second equation).

The third and fourth equations show how we would specify time-varying survival, where survival in a given year is the sum of the logit-mean intercept, mu, plus epsilon\_t, which is a normally-distributed random variable with a mean of zero and a variance of tau\_phi.

And we can do the same thing for recapture probability, in equations 5 and 6.

## Identifiability of the CJS model with time-variation

Now there is something important we need to be aware of: if we make this model fully-time dependent, parameters $\large \phi\_T$ and $\large p\_T$ are not identifiable.

* The model will return posteriors for both parameters (because each has a prior), but the model will **not** be able to separately **estimate both parameters**.
* Posteriors will actually be for

## Identifiability of the CJS model with time-variation 2

**Why is this?** Here is a simple example, a CMR study with two occasions, where we captured 100 individuals in the first occasion and 60 were recaptured in occasion 2.

Well, we can get 60 recaptures with survival of 0.8, and a recapture probability of 0.75, or with higher survival and lower recapture, or with lower survival and higher recapture.

There is no unique solution! We must have more than 2 occasions and internal zeros to have enough information to disentangle survival and detection. So, always design your CJS study to have at least 3 or more sampling occasions.

## Workarounds

Use a more simple model with a constant recapture probability.

Include covariatsion on survival or recapture probabilities.

Or use informed priors.

## Assumptions of the CJS model

1. Every animal has the same chance of capture,
2. Every animal has same probability of surviving
3. Marks are not lost
4. Samples are instantaneous (short periods)
5. All emigration is permanent (101 must indicate )
6. Fates of individuals are independent of others

**Design your study to meet these assumptions**:

* Fixed study area
* Study area is ~large
* Samples are ~instantaneous
* Individuals are randomly sampled

## Analysis in R