

Chapter 7

Matrix Models IV: Integral Projection Models

“If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.”

— John von Neumann

As seen in Chapter 5, MPMs may be estimated using functions representing vital rates, where these vital rate functions may then be used to estimate each matrix element. We term this approach the **function-based MPM**, because of its use of functions to estimate matrix elements. One reason that this approach is very powerful is that it allows population ecologists to create simulations in which vital rates themselves are manipulated, for example via altered climate relationships or management regimes.

Easterling et al. (2000) proposed a special case of the function-based MPM called the **integral projection model (IPM)**. In integral projection models, the familiar projection equation $\mathbf{n}_{t+1} = \mathbf{A}\mathbf{n}_t$ changes to

$$n(k, t + 1) = \int_L^U K(k, j)n(j, t)dj \quad (7.1)$$

where an individual in state j in time t either transitions to state k or produces offspring in state k in time $t + 1$, $n(j, t)dj$ refers to the number of individuals with their state in the range between j and $j + dj$, L and U represent the lower and upper bounds of the state variable, and $K(k, j)$ is the projection kernel $K(k, j) = P(k, j) + F(k, j)$. In this projection kernel, $P(k, j)$ represents the survival-transition probability from state j in time t to state k in time $t + 1$, and $F(k, j)$ represents the production of offspring in state k in time $t + 1$ by an individual in state j in time t . Because this equation is written as an integral as opposed to a discrete summation, it may appear to be quite different from the matrix approaches that we have seen so far. Indeed, there are those who have attempted to use these equations in their pure, integral forms, and when used analytically IPMs are not really MPMs. However, in practice, IPMs are typically discretized so that they may be parameterized as matrices. This discretization allows practitioners to use matrix approaches for analysis, because the projection kernel in equation 7.1 becomes perfectly analogous to the discrete projection equation $\mathbf{n}_{t+1} = \mathbf{A}\mathbf{n}_t$.

Just as in other function-based MPMs, the IPM generally assumes that survival probabilities follow a binomial distribution and so may be estimated via generalized linear models, generalized linear mixed models, general additive models, or related approaches assuming a binomial response. Likewise, probabilities of reproduction, observation, or maturity, should also follow a binomial response. We see a difference in the estimation of the size probability, because IPMs require the use of a continuous size metric. In fact, the “traditional” IPM assumes a Gaussian distribution for size, and the underlying

assumptions of the Gaussian distribution make assumptions that may fail in many circumstances and hence lead to biased results. A more flexible IPM approach allows the use of other distributions, including continuous distributions such as the gamma distribution, and discrete distributions such as the Poisson or negative binomial distributions when necessary. Package `lefk3` allows all of these distributions.

Discretized IPMs assume vital rate models that parameterize the main kernel populating the matrix elements. Let's review the fourteen vital rate models possible in `lefk3`:

1. **Survival probability** - This is the probability of surviving from occasion t to occasion $t+1$, given that the individual is in stage j in occasion t (and, if historical, in stage l in occasion $t-1$). In `lefk3`, this parameter may be modeled as a function of up to three size metrics, reproductive status, patch, year, age, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$. *This parameter is required in all function-based matrices.*
2. **Observation probability** - This is the probability of observation in occasion $t+1$ of an individual in stage k given survival from occasion t to occasion $t+1$. This parameter is only used when at least one stage is technically not observable. For example, some plants are capable of vegetative dormancy, in which case they are alive but do not necessarily sprout in all years. In these cases, the probability of sprouting may be estimated as the observation probability. Note that this probability does not refer to observer effort, and so should only be used to differentiate completely unobservable stages where the observation status refers to an important biological phenomenon, such as when individuals may be alive but have a size of zero. In `lefk3`, this parameter may be modeled as a function of up to three size metrics, reproductive status, patch, year, age, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.
3. **Primary size transition probability** - This is the probability of becoming size k in occasion $t+1$ assuming survival from occasion t to occasion $t+1$ and observation in that time. If multiple size metrics are used, then this refers only to the first of these, which we may refer to as the *primary size variable*. In `lefk3`, this parameter may be modeled as a function of up to three size metrics, reproductive status, patch, year, age, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$. *This parameter is required in all function-based size-classified matrices.*
4. **Secondary size transition probability** - This is the probability of becoming size k in occasion $t+1$ assuming survival from occasion t to occasion $t+1$ and observation in that time, within a second size metric used for classification in addition to the primary metric. In `lefk3`, this parameter may be modeled as a function of up to three size metrics, reproductive status, patch, year, age, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.
5. **Tertiary size transition probability** - This is the probability of becoming size k in occasion $t+1$ assuming survival from occasion t to occasion $t+1$ and observation in that time, within a third size metric used for classification in addition to the primary and secondary metrics. In `lefk3`, this parameter may be modeled as a function of up to three size metrics, reproductive status, patch, year, age, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.
6. **Reproduction probability** - This is the probability of becoming reproductive in occasion $t+1$ given survival from occasion t to occasion $t+1$, and observation in that time. Note that this should be used only if the researcher wishes to separate breeding from non-breeding mature stages. If all adult stages are potentially reproductive and no separation of reproducing from non-reproducing adults is required by the life history model, then this parameter should not be estimated. In `lefk3`, this parameter may be modeled as a function of up to three size metrics, reproductive status, patch, year, age, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.
7. **Fecundity rate** - Under the default setting, this is the rate of successful production of offspring

in occasion t by individuals alive, observable, and reproductive in that time, and, if assuming a pre-breeding model and sufficient information is provided in the dataset, the survival of those offspring into occasion $t+1$ in whatever juvenile class is possible. Thus, the fecundity rate of seed-producing plants might be split into seedlings, which are plants that germinated within a year of seed production, and dormant seeds. Alternatively, it may be given only as produced fruits or seeds, with the survival and germination of seeds provided elsewhere in the MPM development process, such as within a supplement table. An additional setting allows fecundity rate to be estimated using data provided for occasion $t+1$ instead of occasion t . In `lefko3`, this parameter may be modeled as a function of up to three size metrics, reproductive status, patch, year, age, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.

8. **Juvenile survival probability** - This is the probability of surviving from juvenile stage j in occasion t to occasion $t+1$. It is used only when the user wishes to model vital rates for a single size-unclassified juvenile period separately from adults. In `lefko3`, this parameter may be modeled as a function of up to three size metrics, patch, year, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.
9. **Juvenile observation probability** - This is the probability of observation in occasion $t+1$ of an individual in juvenile stage j in occasion t given survival from occasion t to occasion $t+1$. It is used only when the user wishes to model vital rates for a single size-unclassified juvenile period separately from adults. In `lefko3`, this parameter may be modeled as a function of up to three size metrics, patch, year, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.
10. **Juvenile primary size transition probability** - This is the probability of becoming size-classified stage k in occasion $t+1$ assuming survival from juvenile stage j in occasion t to occasion $t+1$ and observation in that time. It is in terms of a single size metric. It is used only when the user wishes to model vital rates for a single size-unclassified juvenile period separately from adults. In `lefko3`, this parameter may be modeled as a function of up to three size metrics, patch, year, and a number of individual or environmental covariates in occasions t and $t-1$, and individual identity.
11. **Juvenile secondary size transition probability** - This is the probability of becoming size-classified stage k in occasion $t+1$ assuming survival from juvenile stage j in occasion t to occasion $t+1$ and observation in that time, in a secondary size metric in addition to the primary size metric. It is used only when the user wishes to model vital rates for a single size-unclassified juvenile period separately from adults. In `lefko3`, this parameter may be modeled as a function of up to three size metrics, patch, year, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.
12. **Juvenile tertiary size transition probability** - This is the probability of becoming size-classified stage k in occasion $t+1$ assuming survival from juvenile stage j in occasion t to occasion $t+1$ and observation in that time, in a tertiary size metric in addition to the primary and secondary size metrics. It is used only when the user wishes to model vital rates for a single size-unclassified juvenile period separately from adults. In `lefko3`, this parameter may be modeled as a function of up to three size metrics, patch, year, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.
13. **Juvenile reproduction probability** - This is the probability of reproducing in mature stage k in occasion $t+1$ given survival from juvenile stage j in occasion t to occasion $t+1$, and observation in that time. It is used only when the user wishes to model vital rates for a single size-unclassified juvenile period separately from adults. In `lefko3`, this parameter may be modeled as a function of up to three size metrics, patch, year, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$.
14. **Juvenile maturity probability** - This is the probability of becoming mature in occasion $t+1$ given survival from juvenile stage j in occasion t to occasion $t+1$. It is used only when the user wishes to model vital rates for a single size-unclassified juvenile period separately from adults.

In `lefk3`, this parameter may be modeled as a function of up to three size metrics, patch, year, and individual identity, and a number of individual or environmental covariates in occasions t and $t-1$. Note that this parameter denotes transition to maturity.

Of these fourteen vital rates, most users will estimate at least parameters (1) survival probability, (3) size transition probability, and (7) fecundity. These three are the default set for function `modelsearch()`. Parameters (2) observation probability and (6) reproduction probability may be used when some stages are included that are completely unobservable (and so do not have any size), or that are mature but non-reproductive, respectively. Parameters (4) secondary size transition and (5) tertiary size transition should only be used when size classification involves more than one size variable. Parameters (8) through (14) should only be added if the dataset contains juvenile individuals transitioning to maturity, and these juveniles live essentially as a single juvenile stage for some amount of time before transitioning to maturity, or before transitioning to a stage that is size-classified in the same manner as adult stages are. If juveniles can be classified by size similarly to adults (or at least on the same scale), then only vital rates (1) through (7) should be used. If multiple juvenile stages exist on a different size classification system than adults, then we advise the user to model only rates (1) through (7) and to use **stage groups** to stratify vital rate models properly.

Let us assume that the state of the individual is represented by a continuous variable, such as a continuous size metric. This continuous size metric will be used to estimate parameter (3), the primary size transition probability. It may be Gaussian distributed, as is often assumed in IPMs (Doak et al., 2021), or may follow a different continuous distribution such as the gamma distribution. The survival-transition kernel $P(k, j)$ and the fecundity kernel $F(k, j)$ will be estimated as in the function-based MPM, as products of conditional rates or probabilities that are themselves estimated via linear models, additive models, or some other function-based approach. How, then, is the primary size transition probability estimated?

In practice, the continuous state variable is broken down into a series of continuous domains, each with its own midpoint and upper and lower bounds. Individual domains under the Gaussian and gamma distributions are shown in figures 7.1 a and d, respectively. The domain midpoints are sometimes referred to as *mesh points*, and together with their upper and lower bounds they compose a series of *size bins* of generally equal size. To approximate a continuous size, we choose a rather high number of size bins, m , perhaps on the order of 100 or even more.

7.1 Midpoint method vs. cumulative density function (CDF)

There are several methods to estimate the size transition probabilities associated with each size bin. The original method developed is referred to as the **midpoint method** (Doak et al., 2021). It is the default in packages such as `IPMpack` and in some published guides to IPM creation (Metcalf et al., 2013; Merow et al., 2014). The mesh points are then defined as $j_i = L + (i - 0.5)h$, where i is the set of integers from 1 to m ($i = 1, 2, \dots, m$), L is the lower state bound as before, and h is the width of the state bin or size bin, given as $h = (U - L)/m$. If each kernel is composed of a vital rate function assuming some sort of probability distribution, and size is distributed on a Gaussian, gamma, or other continuous distribution, then h accounts for the area under the distribution density curve for the corresponding vital rate contributing to the kernel at each midpoint size value used in the model (figures 7.1 b and e). If we think of the integral as being approximated by a series of rectangles under the function being integrated, then h accounts for the width of the rectangle in the area approximation. Thus, we have the following.

$$n(x_j, t + 1) = h \sum_{i=1}^m K(x_j, x_i) n(x_i, t) \quad (7.2)$$

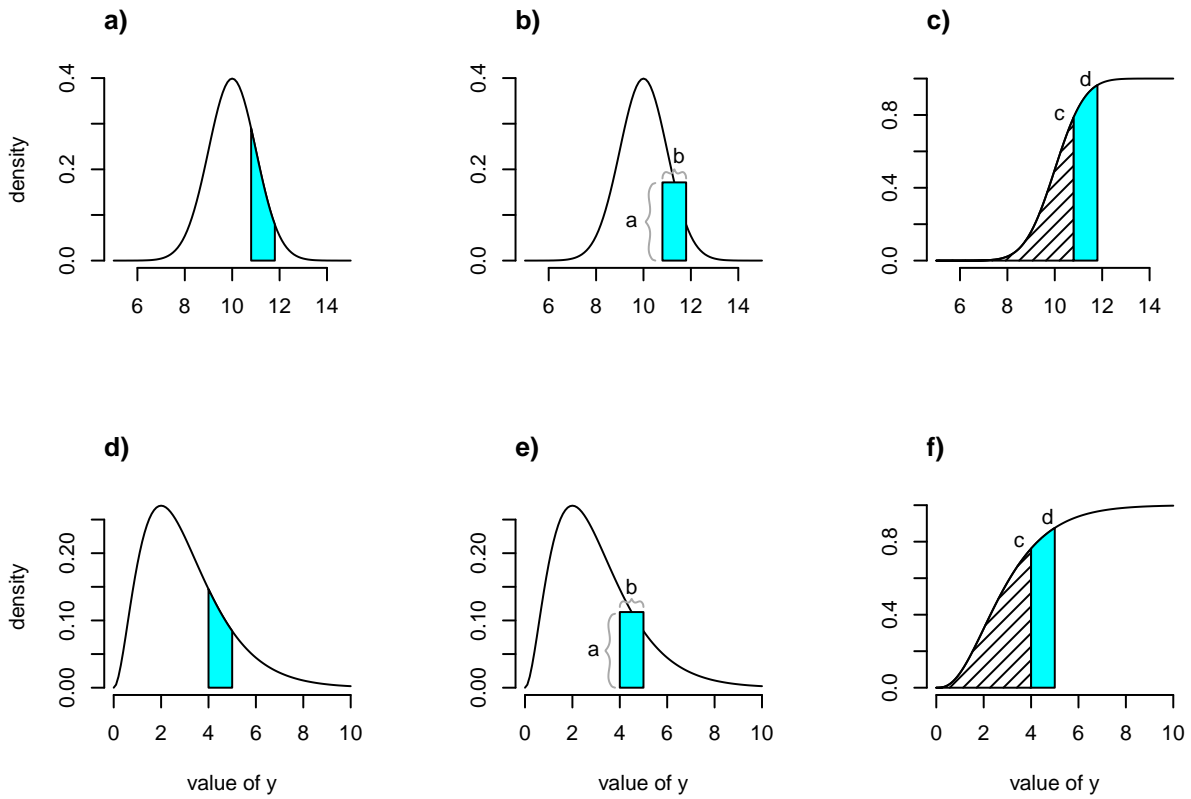


Figure 7.1: Shaded regions indicate the true size transition probability involved in transition (a,d), the estimated size transition probability via the midpoint method (b,e), and the estimated size transition probability via CDF (c,f) for the Gaussian (a-c) and gamma (d-f) distributions.

Doak et al. (2021) pointed out that the midpoint method yields biased results, often overestimating size transition probabilities. They proposed a second method based on the cumulative density function associated with the continuous distribution being used. We will call this the **CDF** method here. In this method, the cumulative probability associated with the lower and higher boundaries of the size bin are first calculated. Then, the cumulative probability associated with the lower boundary is subtracted from the cumulative probability associated with the higher boundary, yielding the exact probability associated with the size bin itself (figures 7.1 c and f). This method does not yield biased results, and so is the default method used in `lefk3` (although the midpoint method is available as an option).

The practical impact is that this approach has us creating size bins, just as in the function-based approach. If we have created size bins, then we have essentially created size-classified life history stages. Thus, equation 7.2 can be treated as a matrix projection, perfectly analogous to $\mathbf{n}_{t+1} = \mathbf{K}\mathbf{n}_t$. So from a practical standpoint, an integral projection model is simply a function-based matrix projection model in which a continuous size metric determines demography. Indeed, Ellner and Rees (2006) further proposed a generalization allowing IPMs to be developed with discrete stages in some portions of the life history, referring to this as a **complex integral projection model**. So in practice, there is now virtually no practical difference between IPMs and function-based MPMs, although there are theoretical differences due to the assumption of integrals over continuous size in the former.

7.2 Creating IPMs in `lefk3`

How do we create IPMs in package `lefk3`? This turns out to be quite easy. To illustrate the process, we will use the *Lathyrus vernus* dataset (see section 1.6.2). In this exercise, we will create both ahistorical and historical IPMs.

First, we will clear memory and take a look at the dataset.

```
rm(list=ls(all=TRUE))

library(lefk3)
data(lathyrus)
summary(lathyrus)
> SUBPLOT GENET Volume88 lnVol188
> Min. :1.000 Min. : 1.0 Min. : 3.4 Min. :1.200
> 1st Qu.:2.000 1st Qu.: 48.0 1st Qu.: 63.0 1st Qu.:4.100
> Median :3.000 Median : 97.0 Median : 732.5 Median :6.600
> Mean :3.223 Mean :110.2 Mean : 749.4 Mean :5.538
> 3rd Qu.:4.000 3rd Qu.:167.5 3rd Qu.:1025.5 3rd Qu.:6.900
> Max. :6.000 Max. :284.0 Max. :7032.0 Max. :8.900
> NA's :404 NA's :404
> FCODE88 Flow88 Intactseed88 Dead1988 Dormant1988
> Min. :0.0000 Min. : 1.00 Min. : 0 Mode:logical Mode:logical
> 1st Qu.:0.0000 1st Qu.: 4.00 1st Qu.: 0 NA's:1119 NA's:1119
> Median :0.0000 Median : 8.00 Median : 0
> Mean :0.3399 Mean :11.86 Mean : 3
> 3rd Qu.:1.0000 3rd Qu.:15.00 3rd Qu.: 4
> Max. :1.0000 Max. :66.00 Max. :34
> NA's :404 NA's :910 NA's :875
> Missing1988 Seedling1988 Volume89 lnVol189
> Mode:logical Min. :1.000 Min. : 1.8 Min. :0.600
> NA's:1119 1st Qu.:2.000 1st Qu.: 15.6 1st Qu.:2.700
> Median :2.000 Median : 118.8 Median :4.800
> Mean :2.144 Mean : 573.3 Mean :4.855
```

```

>          3rd Qu.:3.000   3rd Qu.: 968.8   3rd Qu.:6.900
>          Max.    :3.000   Max.    :6539.4   Max.    :8.800
>          NA's    :1022   NA's    :294     NA's    :294
>   FCODE89      Flow89      Intactseed89      Dead1989
>   Min.    :0.0000   Min.    : 1.00   Min.    : 0.000   Min.    :1
>   1st Qu.:0.0000   1st Qu.: 5.00   1st Qu.: 0.000   1st Qu.:1
>   Median :0.0000   Median :11.00   Median : 5.000   Median :1
>   Mean    :0.2667   Mean    :14.88   Mean    : 8.273   Mean    :1
>   3rd Qu.:1.0000   3rd Qu.:20.00   3rd Qu.:13.000   3rd Qu.:1
>   Max.    :1.0000   Max.    :97.00   Max.    :66.000   Max.    :1
>   NA's    :294     NA's    :906     NA's    :899     NA's    :1077
>   Dormant1989   Missing1989   Seedling1989   Volume90      lnVol190
>   Min.    :1      Min.    :1      Min.    :1.000   Min.    : 2.1   Min.    :0.700
>   1st Qu.:1      1st Qu.:1      1st Qu.:2.000   1st Qu.: 12.6   1st Qu.:2.500
>   Median :1      Median :1      Median :2.000   Median : 61.0   Median :4.100
>   Mean    :1      Mean    :1      Mean    :2.136   Mean    :244.1   Mean    :4.207
>   3rd Qu.:1      3rd Qu.:1      3rd Qu.:2.000   3rd Qu.:295.2   3rd Qu.:5.700
>   Max.    :1      Max.    :1      Max.    :3.000   Max.    :4242.8   Max.    :8.400
>   NA's    :1046   NA's    :1112   NA's    :1001   NA's    :245     NA's    :245
>   FCODE90      Flow90      Intactseed90      Dead1990
>   Min.    :0.0000   Min.    : 1.000   Min.    : 0.000   Min.    :1
>   1st Qu.:0.0000   1st Qu.: 3.000   1st Qu.: 0.000   1st Qu.:1
>   Median :0.0000   Median : 6.000   Median : 0.000   Median :1
>   Mean    :0.1581   Mean    : 8.104   Mean    : 2.514   Mean    :1
>   3rd Qu.:0.0000   3rd Qu.:10.750   3rd Qu.: 1.000   3rd Qu.:1
>   Max.    :1.0000   Max.    :54.000   Max.    :37.000   Max.    :1
>   NA's    :246     NA's    :985     NA's    :981     NA's    :1007
>   Dormant1990   Missing1990   Seedling1990   Volume91      lnVol191
>   Min.    :1      Min.    :1      Min.    :1.000   Min.    : 4.0   Min.    :1.400
>   1st Qu.:1      1st Qu.:1      1st Qu.:2.000   1st Qu.: 12.0   1st Qu.:2.500
>   Median :1      Median :1      Median :2.000   Median : 118.5   Median :4.800
>   Mean    :1      Mean    :1      Mean    :2.186   Mean    :418.7   Mean    :4.642
>   3rd Qu.:1      3rd Qu.:1      3rd Qu.:2.000   3rd Qu.: 689.7   3rd Qu.:6.500
>   Max.    :1      Max.    :1      Max.    :3.000   Max.    :6645.8   Max.    :8.800
>   NA's    :1054   NA's    :1105   NA's    :1049   NA's    :305     NA's    :305
>   FCODE91      Flow91      Intactseed91      Dead1991      Dormant1991
>   Min.    :0.0000   Min.    : 1.00   Min.    : 0.000   Min.    :1     Min.    :1
>   1st Qu.:0.0000   1st Qu.: 4.00   1st Qu.: 0.000   1st Qu.:1     1st Qu.:1
>   Median :0.0000   Median : 8.00   Median : 3.500   Median :1     Median :1
>   Mean    :0.2525   Mean    :11.12   Mean    : 5.805   Mean    :1     Mean    :1
>   3rd Qu.:1.0000   3rd Qu.:15.00   3rd Qu.:10.000   3rd Qu.:1     3rd Qu.:1
>   Max.    :1.0000   Max.    :48.00   Max.    :48.000   Max.    :1     Max.    :1
>   NA's    :307     NA's    :954     NA's    :919     NA's    :925     NA's    :1034
>   Missing1991   Seedling1991
>   Min.    :1      Min.    :1.000
>   1st Qu.:1      1st Qu.:2.000
>   Median :1      Median :2.000
>   Mean    :1      Mean    :1.973
>   3rd Qu.:1      3rd Qu.:2.000
>   Max.    :1      Max.    :3.000
>   NA's    :1095   NA's    :1082

```

This dataset includes information on 1,119 individuals, so there are 1,119 rows with data. There are 38 columns. The first two columns give identifying information about each individual (SUBPLOT refers to the patch, and GENET refers to individual identity), with each individual's data entirely restricted to one row. This is followed by four sets of nine columns, each named `VolumeXX`, `lnVolXX`, `FCODEXX`, `FlowXX`, `IntactseedXX`, `Dead19XX`, `DormantXX`, `Missing19XX`, and `SeedlingXX`, where `XX` corresponds to the year of observation and with years organized consecutively. Thus, columns 3-11 refer to year 1988, columns 12-20 refer to year 1989, etc.

7.2.1 Developing stageframes for IPMs

First, we will create a stageframe for this dataset. We will base our stageframe on the life history model provided in Ehrlén (2000), but use a different size classification based on leaf volume to allow IPM construction and make all mature stages other than vegetative dormancy reproductive (figure 7.2).

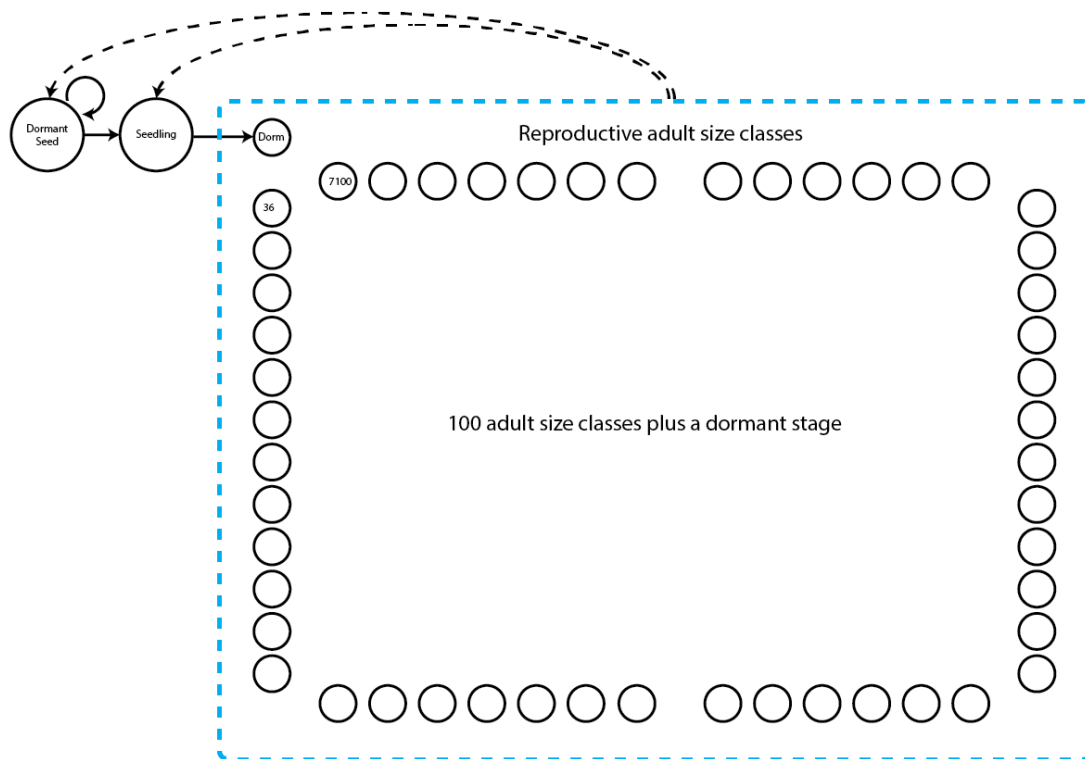


Figure 7.2: Life history model of *Lathyrus vernus*. Not all adult classes are shown. Survival transitions are indicated with solid arrows, while fecundity transitions are indicated with dashed arrows.

In the stageframe code below, we show that we want an IPM by choosing two stages that serve as the size limits for the IPM's discretized size bin classification. **These two size classes should have exactly the same characteristics in the stageframe other than size.** The sizes input into the `sizes` vector for these two stages should not be midpoints. Instead, the size for the lower limit should be the lower limit of the minimum size bin, while the size input for the upper limit should be the upper limit of the maximum size bin. By choosing these two size limits, we can skip adding and describing the many size classes that will fall between these limits - function `sf_create()` will create all of these

for us. We mark these limits in the vector that we load into the `stagenames` option using the string "ipm". We then input all other characteristics for these size bins, such as observation status, maturity status, reproductive status, and these characteristics must be the same for both the minimum and maximum size bins. Package `lefko3` will then create and name all IPM size classes according to its own conventions. The default number of size classes is 100 bins, and this can be altered using the `ipmbins` option. Note that this is essentially the same procedure described in section 2.4.

```
sizevector <- c(0, 100, 0, 1, 7100)
stagevector <- c("Sd", "Sd1", "Dorm", "ipm", "ipm")
repvector <- c(0, 0, 0, 1, 1)
obsvector <- c(0, 1, 0, 1, 1)
matvector <- c(0, 0, 1, 1, 1)
immvector <- c(1, 1, 0, 0, 0)
propvector <- c(1, 0, 0, 0, 0)
indataset <- c(0, 1, 1, 1, 1)
binvec <- c(0, 100, 0.5, 1, 1)
comments <- c("Dormant seed", "Seedling", "Dormant", "ipm adult stage",
  "ipm adult stage")

lathframeipm <- sf_create(sizes = sizevector, stagenames = stagevector,
  repstatus = repvector, obsstatus = obsvector, propstatus = propvector,
  immstatus = immvector, matstatus = matvector, comments = comments,
  indataset = indataset, binhalfwidth = binvec, ipmbins = 100, roundsize = 3)

dim(lathframeipm)
> [1] 103 29
```

This stageframe has 103 stages - dormant seed, seedling, vegetative dormancy, and 100 size-classified adult stages. Let's look at just a few key columns.

```
lathframeipm[,c("stage", "size", "sizebin_min", "sizebin_max", "comments")]
>
>      stage      size sizebin_min sizebin_max      comments
> 1      Sd      0.000         0.00         0.00      Dormant seed
> 2      Sd1    100.000         0.00        200.00      Seedling
> 3      Dorm      0.000        -0.50         0.50      Dormant
> 4  sza_36.495_0  36.495         1.00        71.99 ipm adult stage
> 5  sza_107.48_0 107.485         71.99       142.98 ipm adult stage
> 6  sza_178.47_0 178.475        142.98       213.97 ipm adult stage
> 7  sza_249.46_0 249.465        213.97       284.96 ipm adult stage
> 8  sza_320.45_0 320.455        284.96       355.95 ipm adult stage
> 9  sza_391.44_0 391.445        355.95       426.94 ipm adult stage
> 10 sza_462.43_0 462.435        426.94       497.93 ipm adult stage
> 11 sza_533.42_0 533.425        497.93       568.92 ipm adult stage
> 12 sza_604.41_0 604.415        568.92       639.91 ipm adult stage
> 13 sza_675.40_0 675.405        639.91       710.90 ipm adult stage
> 14 sza_746.39_0 746.395        710.90       781.89 ipm adult stage
> 15 sza_817.38_0 817.385        781.89       852.88 ipm adult stage
> 16 sza_888.37_0 888.375        852.88       923.87 ipm adult stage
> 17 sza_959.36_0 959.365        923.87       994.86 ipm adult stage
> 18 sza_1030.3_0 1030.355       994.86      1065.85 ipm adult stage
> 19 sza_1101.3_0 1101.345      1065.85     1136.84 ipm adult stage
> 20 sza_1172.3_0 1172.335     1136.84     1207.83 ipm adult stage
```

> 21	sza_1243.3_0	1243.325	1207.83	1278.82	ipm adult stage
> 22	sza_1314.3_0	1314.315	1278.82	1349.81	ipm adult stage
> 23	sza_1385.3_0	1385.305	1349.81	1420.80	ipm adult stage
> 24	sza_1456.2_0	1456.295	1420.80	1491.79	ipm adult stage
> 25	sza_1527.2_0	1527.285	1491.79	1562.78	ipm adult stage
> 26	sza_1598.2_0	1598.275	1562.78	1633.77	ipm adult stage
> 27	sza_1669.2_0	1669.265	1633.77	1704.76	ipm adult stage
> 28	sza_1740.2_0	1740.255	1704.76	1775.75	ipm adult stage
> 29	sza_1811.2_0	1811.245	1775.75	1846.74	ipm adult stage
> 30	sza_1882.2_0	1882.235	1846.74	1917.73	ipm adult stage
> 31	sza_1953.2_0	1953.225	1917.73	1988.72	ipm adult stage
> 32	sza_2024.2_0	2024.215	1988.72	2059.71	ipm adult stage
> 33	sza_2095.2_0	2095.205	2059.71	2130.70	ipm adult stage
> 34	sza_2166.1_0	2166.195	2130.70	2201.69	ipm adult stage
> 35	sza_2237.1_0	2237.185	2201.69	2272.68	ipm adult stage
> 36	sza_2308.1_0	2308.175	2272.68	2343.67	ipm adult stage
> 37	sza_2379.1_0	2379.165	2343.67	2414.66	ipm adult stage
> 38	sza_2450.1_0	2450.155	2414.66	2485.65	ipm adult stage
> 39	sza_2521.1_0	2521.145	2485.65	2556.64	ipm adult stage
> 40	sza_2592.1_0	2592.135	2556.64	2627.63	ipm adult stage
> 41	sza_2663.1_0	2663.125	2627.63	2698.62	ipm adult stage
> 42	sza_2734.1_0	2734.115	2698.62	2769.61	ipm adult stage
> 43	sza_2805.1_0	2805.105	2769.61	2840.60	ipm adult stage
> 44	sza_2876.0_0	2876.095	2840.60	2911.59	ipm adult stage
> 45	sza_2947.0_0	2947.085	2911.59	2982.58	ipm adult stage
> 46	sza_3018.0_0	3018.075	2982.58	3053.57	ipm adult stage
> 47	sza_3089.0_0	3089.065	3053.57	3124.56	ipm adult stage
> 48	sza_3160.0_0	3160.055	3124.56	3195.55	ipm adult stage
> 49	sza_3231.0_0	3231.045	3195.55	3266.54	ipm adult stage
> 50	sza_3302.0_0	3302.035	3266.54	3337.53	ipm adult stage
> 51	sza_3373.0_0	3373.025	3337.53	3408.52	ipm adult stage
> 52	sza_3444.0_0	3444.015	3408.52	3479.51	ipm adult stage
> 53	sza_3515.0_0	3515.005	3479.51	3550.50	ipm adult stage
> 54	sza_3585.9_0	3585.995	3550.50	3621.49	ipm adult stage
> 55	sza_3656.9_0	3656.985	3621.49	3692.48	ipm adult stage
> 56	sza_3727.9_0	3727.975	3692.48	3763.47	ipm adult stage
> 57	sza_3798.9_0	3798.965	3763.47	3834.46	ipm adult stage
> 58	sza_3869.9_0	3869.955	3834.46	3905.45	ipm adult stage
> 59	sza_3940.9_0	3940.945	3905.45	3976.44	ipm adult stage
> 60	sza_4011.9_0	4011.935	3976.44	4047.43	ipm adult stage
> 61	sza_4082.9_0	4082.925	4047.43	4118.42	ipm adult stage
> 62	sza_4153.9_0	4153.915	4118.42	4189.41	ipm adult stage
> 63	sza_4224.9_0	4224.905	4189.41	4260.40	ipm adult stage
> 64	sza_4295.8_0	4295.895	4260.40	4331.39	ipm adult stage
> 65	sza_4366.8_0	4366.885	4331.39	4402.38	ipm adult stage
> 66	sza_4437.8_0	4437.875	4402.38	4473.37	ipm adult stage
> 67	sza_4508.8_0	4508.865	4473.37	4544.36	ipm adult stage
> 68	sza_4579.8_0	4579.855	4544.36	4615.35	ipm adult stage
> 69	sza_4650.8_0	4650.845	4615.35	4686.34	ipm adult stage
> 70	sza_4721.8_0	4721.835	4686.34	4757.33	ipm adult stage
> 71	sza_4792.8_0	4792.825	4757.33	4828.32	ipm adult stage

```

> 72  sza_4863.8_0 4863.815 4828.32 4899.31 ipm adult stage
> 73  sza_4934.8_0 4934.805 4899.31 4970.30 ipm adult stage
> 74  sza_5005.7_0 5005.795 4970.30 5041.29 ipm adult stage
> 75  sza_5076.7_0 5076.785 5041.29 5112.28 ipm adult stage
> 76  sza_5147.7_0 5147.775 5112.28 5183.27 ipm adult stage
> 77  sza_5218.7_0 5218.765 5183.27 5254.26 ipm adult stage
> 78  sza_5289.7_0 5289.755 5254.26 5325.25 ipm adult stage
> 79  sza_5360.7_0 5360.745 5325.25 5396.24 ipm adult stage
> 80  sza_5431.7_0 5431.735 5396.24 5467.23 ipm adult stage
> 81  sza_5502.7_0 5502.725 5467.23 5538.22 ipm adult stage
> 82  sza_5573.7_0 5573.715 5538.22 5609.21 ipm adult stage
> 83  sza_5644.7_0 5644.705 5609.21 5680.20 ipm adult stage
> 84  sza_5715.6_0 5715.695 5680.20 5751.19 ipm adult stage
> 85  sza_5786.6_0 5786.685 5751.19 5822.18 ipm adult stage
> 86  sza_5857.6_0 5857.675 5822.18 5893.17 ipm adult stage
> 87  sza_5928.6_0 5928.665 5893.17 5964.16 ipm adult stage
> 88  sza_5999.6_0 5999.655 5964.16 6035.15 ipm adult stage
> 89  sza_6070.6_0 6070.645 6035.15 6106.14 ipm adult stage
> 90  sza_6141.6_0 6141.635 6106.14 6177.13 ipm adult stage
> 91  sza_6212.6_0 6212.625 6177.13 6248.12 ipm adult stage
> 92  sza_6283.6_0 6283.615 6248.12 6319.11 ipm adult stage
> 93  sza_6354.6_0 6354.605 6319.11 6390.10 ipm adult stage
> 94  sza_6425.5_0 6425.595 6390.10 6461.09 ipm adult stage
> 95  sza_6496.5_0 6496.585 6461.09 6532.08 ipm adult stage
> 96  sza_6567.5_0 6567.575 6532.08 6603.07 ipm adult stage
> 97  sza_6638.5_0 6638.565 6603.07 6674.06 ipm adult stage
> 98  sza_6709.5_0 6709.555 6674.06 6745.05 ipm adult stage
> 99  sza_6780.5_0 6780.545 6745.05 6816.04 ipm adult stage
> 100 sza_6851.5_0 6851.535 6816.04 6887.03 ipm adult stage
> 101 sza_6922.5_0 6922.525 6887.03 6958.02 ipm adult stage
> 102 sza_6993.5_0 6993.515 6958.02 7029.01 ipm adult stage
> 103 sza_7064.5_0 7064.505 7029.01 7100.00 ipm adult stage

```

The function `sf_create()` has created our mesh points and associated size bins. This is in addition to the discrete stages covering the dormant seed, seedling, and dormant adult stages. Of course, we could have made this even more complex. For example, we could have created two sets of stages to use as the upper and lower bounds of two sets of continuous size states that differ in some key characteristic, such as reproductive status. We also could have set up the IPM using two or three different size metrics and used the `ipm` option within each or only some of them. This function provides a great deal of flexibility and power to create exactly the life history model that you may want.

7.2.2 Formatting demographic data and testing distribution assumptions

Next, we will format the data into *hfv* format. Because this is an IPM, we need to estimate linear models of vital rates. This will require us either to fix or to remove NAs in size and fecundity, so we will set `NAas0 = TRUE`. We will also set `NRasRep = TRUE` because we will assume that all adult stages other than dormancy are reproductive, and there are mature individuals in the dataset that do not reproduce but need to be included in reproductive stages (setting this option to `TRUE` makes sure that the reproductive status of non-reproductive individuals in potentially reproductive stages is set to 1, although the actual fecundity is not altered). Finally, we will ignore patches marked in the dataset and estimate matrices only for the full population in order to preserve statistical power for vital rate modeling in historical IPM analysis.

In the input to `verticalize3()` below, we utilize a repeating pattern of variable names arranged in the same order for each monitoring occasion. This arrangement allows us to enter only the first variable in each set, as long as `noyears` and `blocksize` are set properly and no gaps or shuffles appear in the dataset. The data management functions that we have created for `lefko3` do not require such repeating patterns, but they do make the required input in the function much shorter and more succinct. Note also that we will use a new individual identity variable that incorporates the patch identity (`indiv_id`), to prevent repeat individual identities across patches.

```
lathyrus$indiv_id <- paste(lathyrus$SUBPLOT, lathyrus$GENET)

lathvertipm <- verticalize3(lathyrus, noyears = 4, firstyear = 1988,
  individcol = "indiv_id", blocksize = 9, juvcol = "Seedling1988",
  sizeacol = "Volume88", repstracol = "FCODE88", fecacol = "Intactseed88",
  deadacol = "Dead1988", nonobsacol = "Dormant1988", stageassign = lathframeipm,
  stagesize = "sizea", censorcol = "Missing1988", censorkeep = NA,
  censorRepeat = TRUE, censor = TRUE, NAas0 = TRUE, NRasRep = TRUE)

summary_hfv(lathvertipm)
>
> This hfv dataset contains 2527 rows, 42 variables, 1 population,
> 1 patch, 1053 individuals, and 3 time steps.
>   rowid      popid  patchid  individ      year2
> Min.   : 1.0    :2527    :2527  Length:2527   Min.   :1988
> 1st Qu.: 237.0          Class :character 1st Qu.:1988
> Median : 522.0          Mode  :character Median :1989
> Mean   : 537.3                    Mean  :1989
> 3rd Qu.: 820.5                    3rd Qu.:1990
> Max.   :1118.0                   Max.   :1990
>   firstseen  lastseen  obsage  obslifespan  sizea1
> Min.   : 0    Min.   : 0    Min.   :1.000  Min.   :0.000  Min.   : 0.0
> 1st Qu.:1988 1st Qu.:1991 1st Qu.:1.000 1st Qu.:2.000 1st Qu.: 0.0
> Median :1988 Median :1991  Median :2.000 Median :3.000 Median : 9.0
> Mean   :1979 Mean   :1981  Mean   :1.822 Mean   :2.437 Mean   :387.3
> 3rd Qu.:1988 3rd Qu.:1991 3rd Qu.:2.000 3rd Qu.:3.000 3rd Qu.:624.6
> Max.   :1990 Max.   :1991  Max.   :3.000 Max.   :3.000 Max.   :7032.0
>   repstral      fecal      juvgiven1      obsstatus1
> Min.   :0.0000  Min.   : 0.0000  Min.   :0.00000  Min.   :0.0000
> 1st Qu.:0.0000 1st Qu.: 0.0000 1st Qu.:0.00000 1st Qu.:0.0000
> Median :0.0000 Median : 0.0000  Median :0.00000  Median :1.0000
> Mean   :0.1805 Mean   : 0.9889  Mean   :0.06292  Mean   :0.5548
> 3rd Qu.:0.0000 3rd Qu.: 0.0000 3rd Qu.:0.00000 3rd Qu.:1.0000
> Max.   :1.0000 Max.   :66.0000  Max.   :1.00000  Max.   :1.0000
>   repstatus1  fecstatus1  matstatus1  alive1
> Min.   :0.0000  Min.   :0.00000  Min.   :0.0000  Min.   :0.0000
> 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000
> Median :0.0000 Median :0.00000  Median :1.0000  Median :1.0000
> Mean   :0.1805 Mean   :0.08983  Mean   :0.5204  Mean   :0.5833
> 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:1.0000
> Max.   :1.0000 Max.   :1.00000  Max.   :1.0000  Max.   :1.0000
>   stage1      stageindex      sizea2      repstra2
> Length:2527  Min.   : 0.000  Min.   : 0.0  Min.   :0.000
> Class :character 1st Qu.: 0.000 1st Qu.: 12.6 1st Qu.:0.000
```

```

> Mode :character      Median :  3.000      Median : 105.6      Median :0.000
>                               Mean  :  7.405      Mean  : 480.8      Mean  :0.237
>                               3rd Qu.: 12.000      3rd Qu.: 732.5      3rd Qu.:0.000
>                               Max.   :103.000      Max.   :7032.0      Max.   :1.000
>   fecu2      juvgiven2      obsstatus2      repstatus2
> Min.   : 0.00      Min.   :0.0000      Min.   :0.0000      Min.   :0.000
> 1st Qu.: 0.00      1st Qu.:0.0000      1st Qu.:1.0000      1st Qu.:0.000
> Median : 0.00      Median :0.0000      Median :1.0000      Median :0.000
> Mean   : 1.14      Mean   :0.1112      Mean   :0.9454      Mean   :0.237
> 3rd Qu.: 0.00      3rd Qu.:0.0000      3rd Qu.:1.0000      3rd Qu.:0.000
> Max.   :66.00      Max.   :1.0000      Max.   :1.0000      Max.   :1.000
>   fecstatus2      matstatus2      alive2      stage2
> Min.   :0.0000      Min.   :0.0000      Min.   :1      Length:2527
> 1st Qu.:0.0000      1st Qu.:1.0000      1st Qu.:1      Class :character
> Median :0.0000      Median :1.0000      Median :1      Mode  :character
> Mean   :0.1053      Mean   :0.8888      Mean   :1
> 3rd Qu.:0.0000      3rd Qu.:1.0000      3rd Qu.:1
> Max.   :1.0000      Max.   :1.0000      Max.   :1
>   stage2index      sizea3      repstra3      fecu3      juvgiven3
> Min.   : 2.00      Min.   : 0.0      Min.   :0.000      Min.   : 0.00      Min.   :0
> 1st Qu.: 4.00      1st Qu.: 10.0      1st Qu.:0.000      1st Qu.: 0.00      1st Qu.:0
> Median : 5.00      Median : 72.0      Median :0.000      Median : 0.00      Median :0
> Mean   : 10.12      Mean   : 389.7      Mean   :0.218      Mean   : 1.29      Mean   :0
> 3rd Qu.: 14.00      3rd Qu.: 543.4      3rd Qu.:0.000      3rd Qu.: 0.00      3rd Qu.:0
> Max.   :103.00      Max.   :6645.8      Max.   :1.000      Max.   :66.00      Max.   :0
>   obsstatus3      repstatus3      fecstatus3      matstatus3
> Min.   :0.0000      Min.   :0.000      Min.   :0.0000      Min.   :1
> 1st Qu.:1.0000      1st Qu.:0.000      1st Qu.:0.0000      1st Qu.:1
> Median :1.0000      Median :0.000      Median :0.0000      Median :1
> Mean   :0.8346      Mean   :0.218      Mean   :0.1116      Mean   :1
> 3rd Qu.:1.0000      3rd Qu.:0.000      3rd Qu.:0.0000      3rd Qu.:1
> Max.   :1.0000      Max.   :1.000      Max.   :1.0000      Max.   :1
>   alive3      stage3      stage3index
> Min.   :0.0000      Length:2527      Min.   : 0.000
> 1st Qu.:1.0000      Class :character      1st Qu.: 4.000
> Median :1.0000      Mode  :character      Median : 5.000
> Mean   :0.9224      Mean   : 8.735
> 3rd Qu.:1.0000      3rd Qu.:11.000
> Max.   :1.0000      Max.   :97.000

```

Before we move on to the next steps in analysis, let's take a closer look at fecundity. In this dataset, fecundity is mostly a count of intact seeds, and only differs in six cases where the seed output was estimated based on other models. To see this, try the following code.

```

writeLines(paste0("Length of fecundity in t+1: ", length(lathvertipm$feca3)))
> Length of fecundity in t+1: 2527
writeLines(paste0("Total non-integer entries in fecundity in occasion t+1: ",
  length(which(lathvertipm$feca3 != round(lathvertipm$feca3))))))
> Total non-integer entries in fecundity in occasion t+1: 0
writeLines(paste0("\nLength of fecundity in t: ", length(lathvertipm$feca2)))
>
> Length of fecundity in t: 2527

```

```

writeLines(paste0("Total non-integer entries in fecundity in occasion t: ",
  length(which(lathvertipm$feca2 != round(lathvertipm$feca2))))))
> Total non-integer entries in fecundity in occasion t: 6
writeLines(paste0("\nLength of fecundity in t-1: ", length(lathvertipm$feca1)))
>
> Length of fecundity in t-1: 2527
writeLines(paste0("Total non-integer entries in fecundity in occasion t-1: ",
  length(which(lathvertipm$feca1 != round(lathvertipm$feca1))))))
> Total non-integer entries in fecundity in occasion t-1: 6

```

We see that we have quite a bit of fecundity data, and that it is overwhelmingly but not exclusively integer. So, we can either treat fecundity as a continuous variable, or round the values and treat fecundity as a count variable. We will choose the latter approach in this analysis.

```

lathvertipm$feca3 <- round(lathvertipm$feca3)
lathvertipm$feca2 <- round(lathvertipm$feca2)
lathvertipm$feca1 <- round(lathvertipm$feca1)

```

Let's now look at size. Ideally, we would assume the Gaussian distribution for this continuous variable. However, strong skew might recommend the gamma distribution. Let's view a density plot (figure 7.3).

```
plot(density(lathvertipm$sizea2))
```

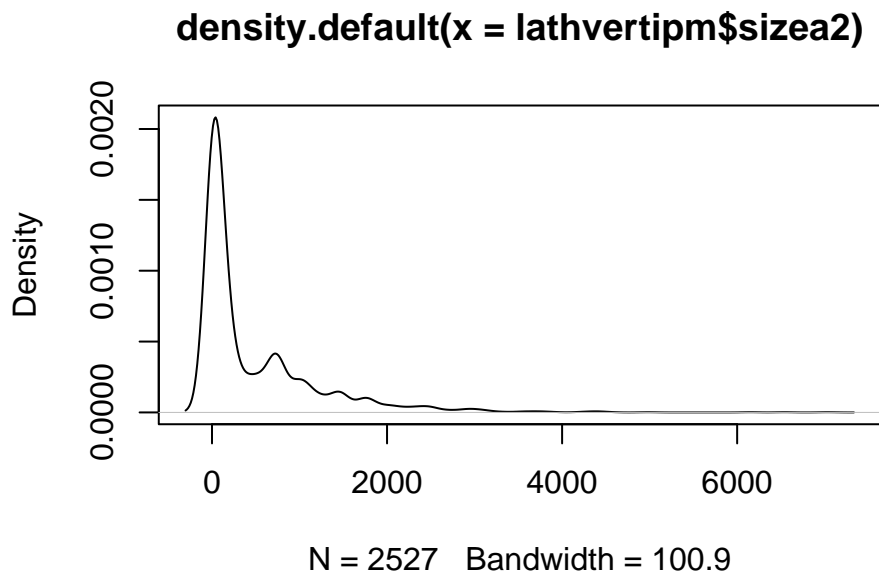


Figure 7.3: Probability density associated with primary size, unaltered

This distribution appears to be arguably right-skewed, but the mean appears to be near the boundary value of zero. We have absorbed the size of zero into the Dormant stage, but there might be reason to avoid using the gamma to avoid issues with zeros. So, we will try using the Gaussian distribution here.

Although we wish to treat fecundity as a count, it is still not clear what underlying distribution we should use. This package currently allows eight choices: Gaussian, gamma, Poisson, negative binomial,

zero-inflated Poisson, zero-inflated negative binomial, zero-truncated Poisson, and zero-truncated negative binomial. To assess which to use, we should first assess whether the mean and variance of the count are equal using a dispersion test. The Poisson distribution assumes that the mean and variance are equal, and so we can test this assumption using a chi-squared test. If it is not significantly different, then we may use some variant of the Poisson distribution. If the data are significantly over- or under-dispersed, then we should use the negative binomial distribution. If fecundity of zero is possible in reproductive stages, as in cases where reproductive status is defined by flowering rather than by offspring production, then we should also test whether the number of zeros is significantly greater than expected under these distributions, and use a zero-inflated distribution if so (if fecundity does not equal zero in any reproductive individuals at all, then we should use a zero-truncated distribution).

Let's look at a plot of the distribution of fecundity (figure 7.4).

```
hist(subset(lathvertipm, repstatus2 == 1)$feca2, main = "Fecundity",
     xlab = "Intact seeds produced in occasion t")
```

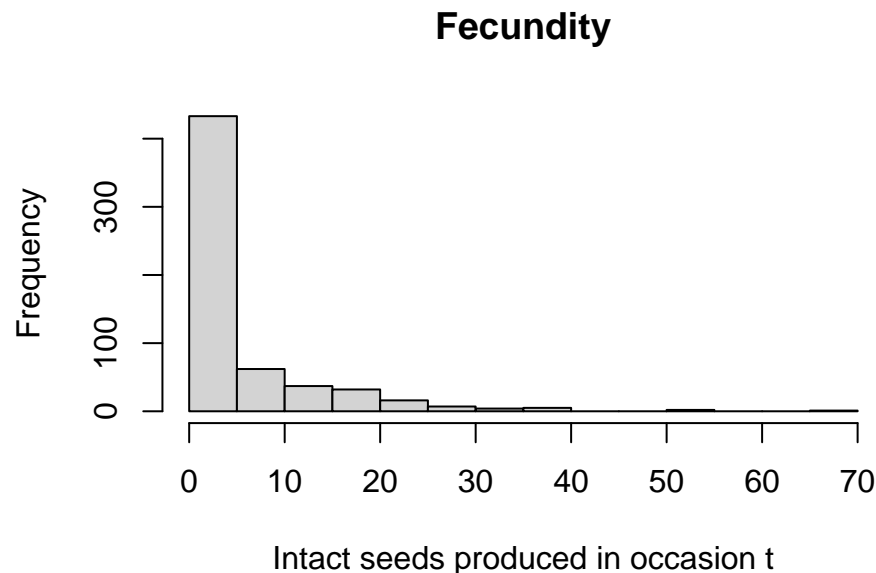


Figure 7.4: Histogram of fecundity

We see that the distribution seems to conform to a classic count variable with a very low mean value. The first bar suggests that there may be too many zeros to use a standard Poisson or negative binomial distribution. But to make that decision, let's formally test the assumptions that the mean equals the variance, and that there are not excess zeros. Both tests use chi-squared distribution-based approaches, with the zero-inflation test based on van den Broek (1995). This is done automatically via the `hfv_qc()` function.

```
hfv_qc(lathvertipm, vitalrates = c("surv", "obs", "size", "fec"),
     juvestimate = "Sd1", indiv = "individ", year = "year2")
> Survival:
>
> Data subset has 43 variables and 2246 transitions.
>
> Variable alive3 has 0 missing values.
> Variable alive3 is a binomial variable.
```

```
>
>
> Observation status:
>
> Data subset has 43 variables and 2121 transitions.
>
> Variable obsstatus3 has 0 missing values.
> Variable obsstatus3 is a binomial variable.
>
>
> Primary size:
>
> Data subset has 43 variables and 1916 transitions.
>
> Variable sizea3 has 0 missing values.
> Variable sizea3 appears to be a floating point variable.
> 1256 elements are not integers.
> The minimum value of sizea3 is 3.4 and the maximum is 6646.
> The mean value of sizea3 is 512.8 and the variance is 507200.
> The value of the Shapiro-Wilk test of normality is 0.7134 with P = 3.014e-49.
> Variable sizea3 differs significantly from a Gaussian distribution.
>
> Variable sizea3 is fully positive, lacking even 0s.
>
>
> Reproductive status:
>
> Data subset has 43 variables and 1916 transitions.
>
> Variable repstatus3 has 0 missing values.
> Variable repstatus3 is a binomial variable.
>
>
> Fecundity:
>
> Data subset has 43 variables and 2246 transitions.
>
> Variable fecal2 has 0 missing values.
> Variable fecal2 appears to be an integer variable.
>
> Variable fecal2 is fully non-negative.
>
> Overdispersion test:
> Mean fecal2 is 1.282
> The variance in fecal2 is 23.21
> The probability of this dispersion level by chance assuming that
> the true mean fecal2 = variance in fecal2,
> and an alternative hypothesis of overdispersion, is 0
> Variable fecal2 is significantly overdispersed.
>
> Zero-inflation and truncation tests:
```



```
> Mean lambda in fecal2 is 0.2774
> The actual number of 0s in fecal2 is 1980
> The expected number of 0s in fecal2 under the null hypothesis is 623
> The probability of this deviation in 0s from expectation by chance is 0
> Variable fecal2 is significantly zero-inflated.
>
>
> Juvenile survival:
>
> Data subset has 43 variables and 281 transitions.
>
> Variable alive3 has 0 missing values.
> Variable alive3 is a binomial variable.
>
>
> Juvenile observation status:
>
> Data subset has 43 variables and 210 transitions.
>
> Variable obsstatus3 has 0 missing values.
> Variable obsstatus3 is a binomial variable.
>
>
> Juvenile primary size:
>
> Data subset has 43 variables and 193 transitions.
>
> Variable sizea3 has 0 missing values.
> Variable sizea3 appears to be a floating point variable.
> 127 elements are not integers.
> The minimum value of sizea3 is 2.1 and the maximum is 61.
> The mean value of sizea3 is 11.23 and the variance is 50.81.
> The value of the Shapiro-Wilk test of normality is 0.5997 with P = 5.72e-21.
> Variable sizea3 differs significantly from a Gaussian distribution.
>
> Variable sizea3 is fully positive, lacking even 0s.
>
>
> Juvenile reproductive status:
>
> Data subset has 43 variables and 193 transitions.
>
> Variable repstatus3 has 0 missing values.
> Variable repstatus3 is a binomial variable.
>
>
> Juvenile maturity status:
>
> Data subset has 43 variables and 210 transitions.
>
> Variable matstatus3 has 0 missing values.
```

```
> Variable matstatus3 is a binomial variable.
```

Such significant results for both tests show us that we should use a zero-inflated negative binomial distribution for fecundity.

Now we will create supplement tables to provide extra data for matrix estimation that is not included in the main demographic dataset. Specifically, we will provide the seed dormancy probability and germination rate, which are given as transitions from the dormant seed stage to another year of seed dormancy or to the germinated seedling stage, respectively. We assume that the germination rate is the same regardless of whether seed was produced in the previous year or has been in the seedbank for longer. We will incorporate these terms both as fixed constants for specific transitions within the resulting matrices, and as multipliers for fecundity, since ultimately fecundity will be estimated as the production of seed multiplied by the seed germination rate or the seed dormancy rate. Because some individuals stay in the seedling stage for only one year, and the seed stage itself cannot be observed and so does not exist in the dataset, we will also set a proxy set of transitions so that R assumes that the transitions from seed in occasion $t-1$ to seedling in occasion t to all mature stages in occasion $t+1$ are equal to the equivalent transitions from seedling in both occasions $t-1$ and t .

We will start with the ahistorical case, and then move on to the historical case, where we also need to input the corresponding stages in occasion $t-1$ and transition types from occasion $t-1$ to t for each transition. Note the use of the "rep", "mat", and "npr" designations in `Stage1` - these are abbreviations telling R to use all reproductive stages, all mature stages, or all non-propagule stages (mature stages plus the seedling stage) in general, respectively.

```
lathsupp2 <- supplemental(stage3 = c("Sd", "Sd1", "Sd", "Sd1"),
  stage2 = c("Sd", "Sd", "rep", "rep"),
  givenrate = c(0.345, 0.054, NA, NA),
  multiplier = c(NA, NA, 0.345, 0.054),
  type = c(1, 1, 3, 3), stageframe = lathframeipm, historical = FALSE)

lathsupp3 <- supplemental(stage3 = c("Sd","Sd","Sd1","Sd1","npr","Sd","Sd1"),
  stage2 = c("Sd", "Sd", "Sd", "Sd", "Sd1", "rep", "rep"),
  stage1 = c("Sd", "rep", "Sd", "rep", "Sd", "mat", "mat"),
  eststage3 = c(NA, NA, NA, NA, "npr", NA, NA),
  eststage2 = c(NA, NA, NA, NA, "Sd1", NA, NA),
  eststage1 = c(NA, NA, NA, NA, "Sd1", NA, NA),
  givenrate = c(0.345, 0.345, 0.054, 0.054, NA, NA, NA),
  multiplier = c(NA, NA, NA, NA, NA, 0.345, 0.054),
  type = c(1, 1, 1, 1, 1, 3, 3), type_t12 = c(1, 2, 1, 2, 1, 1, 1),
  stageframe = lathframeipm, historical = TRUE)

lathsupp2
> stage3 stage2 stage1 eststage3 eststage2 eststage1 givenrate multiplier
> 1 Sd Sd <NA> <NA> <NA> <NA> 0.345 1.000
> 2 Sd1 Sd <NA> <NA> <NA> <NA> 0.054 1.000
> 3 Sd rep <NA> <NA> <NA> <NA> NA 0.345
> 4 Sd1 rep <NA> <NA> <NA> <NA> NA 0.054
> convtype convtype_t12
> 1 1 1
> 2 1 1
> 3 3 1
> 4 3 1

lathsupp3
> stage3 stage2 stage1 eststage3 eststage2 eststage1 givenrate multiplier
```

```

> 1   Sd   Sd   Sd   <NA>   <NA>   <NA>   0.345   1.000
> 2   Sd   Sd   rep  <NA>   <NA>   <NA>   0.345   1.000
> 3   Sd1  Sd   Sd   <NA>   <NA>   <NA>   0.054   1.000
> 4   Sd1  Sd   rep  <NA>   <NA>   <NA>   0.054   1.000
> 5   npr  Sd1  Sd   npr    Sd1   Sd1    NA     1.000
> 6   Sd   rep  mat  <NA>   <NA>   <NA>   NA     0.345
> 7   Sd1  rep  mat  <NA>   <NA>   <NA>   NA     0.054
>   convtype convtype_t12
> 1         1           1
> 2         1           2
> 3         1           1
> 4         1           2
> 5         1           1
> 6         3           1
> 7         3           1

```

7.2.3 Estimating vital rate models

Integral projection models require functions of vital rates to populate them. Here, we will develop these functions as linear models using `modelsearch()`. First, we will create the historical models to assess whether history is a significant influence on vital rates. Note that we have set the appropriate size and fecundity distributions through the settings `sizedist = "gamma"`, `fecdist = "negbin"`, and `fec.zero = TRUE`. We have also set `suite = "size"`, because we have not stratified our size classes by reproductive status.

```

lathmodels3ipm <- modelsearch(lathvertipm, historical = TRUE, approach= "mixed",
  suite = "size", vitalrates = c("surv", "obs", "size", "fec"),
  juvestimate = "Sd1", bestfit = "AICc&k", sizedist = "Gaussian",
  fecdist = "negbin", fec.zero = TRUE, indiv = "individ", year = "year2",
  year.as.random = TRUE, juvsize = TRUE, show.model.tables = TRUE, quiet = TRUE)

```

Now let's see a summary.

```

summary(lathmodels3ipm)
> This LefkoMod object includes 7 linear models.
> Best-fit model criterion used: aicc&k
>
>
>
> Survival model:
>
> Call: stats::glm(formula = alive3 ~ sizea1 + sizea2 + sizea1:sizea2 +
  1, family = "binomial", data = subdata)
>
> Coefficients:
> (Intercept)      sizea1      sizea2 sizea1:sizea2
>  2.161e+00    1.704e-03    1.145e-03   -4.636e-07
>
> Degrees of Freedom: 2245 Total (i.e. Null); 2242 Residual
> Null Deviance:      965.1
> Residual Deviance: 892.4 AIC: 900.4

```

```

>
>
>
> Observation model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: obsstatus3 ~ (1 | year2)
> Data: subdata
>      AIC      BIC    logLik deviance df.resid
> 1351.5346 1362.8539 -673.7673 1347.5346    2119
> Random effects:
> Groups Name          Std.Dev.
> year2 (Intercept) 0
> Number of obs: 2121, groups: year2, 3
> Fixed Effects:
> (Intercept)
>      2.235
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Size model:
> Linear mixed model fit by REML ['lmerMod']
> Formula: sizea3 ~ sizea1 + sizea2 + (1 | year2) + (1 | individ) + sizea1:sizea2
> Data: subdata
> REML criterion at convergence: 29132.25
> Random effects:
> Groups Name          Std.Dev.
> individ (Intercept) 0.0
> year2 (Intercept) 247.2
> Residual 480.4
> Number of obs: 1916, groups: individ, 845; year2, 3
> Fixed Effects:
> (Intercept)          sizea1          sizea2 sizea1:sizea2
> 8.998e+01 3.119e-01 5.954e-01 -9.417e-05
> fit warnings:
> Some predictor variables are on very different scales: consider rescaling
> optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Secondary size model:
> [1] 1
>
>
>
> Tertiary size model:
> [1] 1
>
>

```

```

>
> Reproductive status model:
> [1] 1
>
>
>
> Fecundity model:
> Formula:      feca2 ~ sizea2 + (1 | year2) + (1 | individ)
> Zero inflation: ~sizea1 + sizea2 + (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC    logLik  df.resid
> 2895.248 2952.417 -1437.624    2236
> Random-effects (co)variances:
>
> Conditional model:
> Groups Name      Std.Dev.
> year2  (Intercept) 0.1241
> individ (Intercept) 0.3508
>
> Zero-inflation model:
> Groups Name      Std.Dev.
> year2  (Intercept) 0.3979
> individ (Intercept) 1.0249
>
> Number of obs: 2246 / Conditional model: year2, 3; individ, 931 / Zero-inflation model: year2, 3; i
>
> Dispersion parameter for nbinom2 family (): 2.42
>
> Fixed Effects:
>
> Conditional model:
> (Intercept)      sizea2
> 1.6982784    0.0003123
>
> Zero-inflation model:
> (Intercept)      sizea1      sizea2
> 4.1155831  -0.0004907  -0.0017048
>
>
> Juvenile survival model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: alive3 ~ (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC    logLik  deviance  df.resid
> 323.6696 334.5847 -158.8348 317.6696    278
> Random effects:
> Groups Name      Std.Dev.
> individ (Intercept) 0.0003658
> year2  (Intercept) 0.0000000

```

```

> Number of obs: 281, groups:  individ, 281; year2, 3
> Fixed Effects:
> (Intercept)
>      1.084
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Juvenile observation model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: obsstatus3 ~ sizea2 + (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC  logLik deviance df.resid
> 61.3733 74.7617 -26.6867  53.3733     206
> Random effects:
> Groups Name          Std.Dev.
> individ (Intercept) 53.003
> year2   (Intercept)  1.206
> Number of obs: 210, groups:  individ, 210; year2, 3
> Fixed Effects:
> (Intercept)      sizea2
>  12.83279      0.03985
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Juvenile size model:
> Linear mixed model fit by REML ['lmerMod']
> Formula: sizea3 ~ sizea2 + (1 | year2)
> Data: subdata
> REML criterion at convergence: 1243.682
> Random effects:
> Groups Name          Std.Dev.
> year2   (Intercept)  1.955
> Residual                    5.995
> Number of obs: 193, groups:  year2, 3
> Fixed Effects:
> (Intercept)      sizea2
>   3.0848      0.8465
>
>
> Juvenile secondary size model:
> [1] 1
>
>
> Juvenile tertiary size model:
> [1] 1

```

```

>
>
>
> Juvenile reproduction model:
> [1] 1
>
>
>
> Juvenile maturity model:
> [1] 1
>
>
>
> Number of models in survival table: 5
>
> Number of models in observation table: 5
>
> Number of models in size table: 5
>
> Number of models in secondary size table: 1
>
> Number of models in tertiary size table: 1
>
> Number of models in reproduction status table: 1
>
> Number of models in fecundity table: 25
>
> Number of models in juvenile survival table: 2
>
> Number of models in juvenile observation table: 2
>
> Number of models in juvenile size table: 2
>
> Number of models in juvenile secondary size table: 1
>
> Number of models in juvenile tertiary size table: 1
>
> Number of models in juvenile reproduction table: 1
>
> Number of models in juvenile maturity table: 1
>
>
>
>
> General model parameter names (column 1), and
> specific names used in these models (column 2):
>           parameter_names mainparams
> 1           time t           year2

```

```

> 2             individual  individ
> 3             patch      patch
> 4     alive in time t+1  surv3
> 5     observed in time t+1  obs3
> 6     sizea in time t+1    size3
> 7     sizeb in time t+1    sizeb3
> 8     sizec in time t+1    sizec3
> 9     reproductive status in time t+1  repst3
> 10    fecundity in time t+1    fec3
> 11    fecundity in time t      fec2
> 12    sizea in time t          size2
> 13    sizea in time t-1        size1
> 14    sizeb in time t          sizeb2
> 15    sizeb in time t-1        sizeb1
> 16    sizec in time t          sizec2
> 17    sizec in time t-1        sizec1
> 18    reproductive status in time t    repst2
> 19    reproductive status in time t-1  repst1
> 20    maturity status in time t+1      matst3
> 21    maturity status in time t        matst2
> 22    age in time t                    age
> 23    density in time t                density
> 24    individual covariate a in time t  indcova2
> 25    individual covariate a in time t-1  indcova1
> 26    individual covariate b in time t    indcovb2
> 27    individual covariate b in time t-1  indcovb1
> 28    individual covariate c in time t    indcovc2
> 29    individual covariate c in time t-1  indcovc1
> 30    stage group in time t              group2
> 31    stage group in time t-1            group1
>
>
>
>
> Quality control:
>
> Survival estimated with 931 individuals and 2246 individual transitions.
> Survival accuracy is 0.944.
> Observation estimated with 858 individuals and 2121 individual transitions.
> Observation accuracy is 0.903.
> Primary size estimated with 845 individuals and 1916 individual transitions.
> Primary size pseudo R-squared is 0.571.
> Secondary size transition not estimated.
> Tertiary size transition not estimated.
> Reproduction probability not estimated.
> Fecundity estimated with 931 individuals and 2246 individual transitions.
> Fecundity pseudo R-squared is 0.271.
> Juvenile survival estimated with 281 individuals and 281 individual transitions.
> Juvenile survival accuracy is 0.747.
> Juvenile observation estimated with 210 individuals and 210 individual transitions.

```



```

> Juvenile observation accuracy is 1.
> Juvenile primary size estimated with 193 individuals and 193 individual transitions.
> Juvenile primary size pseudo R-squared is 0.332.
> Juvenile secondary size transition not estimated.
> Juvenile tertiary size transition not estimated.
> Juvenile reproduction probability not estimated.
> Juvenile maturity probability not estimated.

```

We see the influence of history on survival and size and fecundity. So, the historical IPM is the correct choice here. However, we will also create an ahistorical IPM for comparison. For that purpose, we will create the ahistorical vital rate model set.

```

lathmodels2ipm <- modelsearch(lathvertipm, historical = FALSE,
  approach = "mixed", suite = "size",
  vitalrates = c("surv", "obs", "size", "fec"), juvestimate = "Sd1",
  bestfit = "AICc&k", sizedist = "gaussian", fecdist = "negbin",
  fec.zero = TRUE, indiv = "individ", year = "year2", year.as.random = TRUE,
  juvsize = TRUE, show.model.tables = TRUE, quiet = TRUE)

```

Let's see a summary.

```

summary(lathmodels2ipm)
> This LefkoMod object includes 7 linear models.
> Best-fit model criterion used: aicc&k
>
>
>
> Survival model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: alive3 ~ (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC   logLik  deviance  df.resid
> 709.9383 727.0890 -351.9691  703.9383    2243
> Random effects:
> Groups Name      Std.Dev.
> individ (Intercept) 14.6
> year2 (Intercept) 0.0
> Number of obs: 2246, groups:  individ, 931; year2, 3
> Fixed Effects:
> (Intercept)
>      10.03
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
>
> Observation model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: obsstatus3 ~ (1 | year2) + (1 | individ)

```

```

> Data: subdata
>      AIC      BIC    logLik deviance df.resid
> 1337.3193 1354.2982 -665.6596 1331.3193    2118
> Random effects:
> Groups Name      Std.Dev.
> individ (Intercept) 1.216
> year2   (Intercept) 0.000
> Number of obs: 2121, groups:  individ, 858; year2, 3
> Fixed Effects:
> (Intercept)
>      2.788
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Size model:
> Linear mixed model fit by REML ['lmerMod']
> Formula: sizea3 ~ sizea2 + (1 | year2) + (1 | individ)
> Data: subdata
> REML criterion at convergence: 29294.15
> Random effects:
> Groups Name      Std.Dev.
> individ (Intercept) 0.0
> year2   (Intercept) 210.9
> Residual                504.6
> Number of obs: 1916, groups:  individ, 845; year2, 3
> Fixed Effects:
> (Intercept)      sizea2
>      164.0695      0.6211
> optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Secondary size model:
> [1] 1
>
>
>
> Tertiary size model:
> [1] 1
>
>
>
> Reproductive status model:
> [1] 1
>
>
>
> Fecundity model:
> Formula:      fecu2 ~ sizea2 + (1 | year2) + (1 | individ)
> Zero inflation: ~sizea2 + (1 | year2) + (1 | individ)

```

```

> Data: subdata
>      AIC      BIC    logLik  df.resid
> 2905.970 2957.422 -1443.985    2237
> Random-effects (co)variances:
>
> Conditional model:
> Groups Name      Std.Dev.
> year2  (Intercept) 0.1432
> individ (Intercept) 0.3305
>
> Zero-inflation model:
> Groups Name      Std.Dev.
> year2  (Intercept) 0.3898
> individ (Intercept) 1.0047
>
> Number of obs: 2246 / Conditional model: year2, 3; individ, 931 / Zero-inflation model: year2, 3; i
>
> Dispersion parameter for nbinom2 family (): 2.17
>
> Fixed Effects:
>
> Conditional model:
> (Intercept)      sizea2
> 1.736911      0.000286
>
> Zero-inflation model:
> (Intercept)      sizea2
> 4.014336      -0.001969
>
>
> Juvenile survival model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: alive3 ~ (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC    logLik  deviance  df.resid
> 323.6696 334.5847 -158.8348  317.6696    278
> Random effects:
> Groups Name      Std.Dev.
> individ (Intercept) 0.0003658
> year2  (Intercept) 0.0000000
> Number of obs: 281, groups:  individ, 281; year2, 3
> Fixed Effects:
> (Intercept)
> 1.084
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Juvenile observation model:

```

```

> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: obsstatus3 ~ sizea2 + (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC  logLik deviance df.resid
> 61.3733  74.7617 -26.6867  53.3733     206
> Random effects:
> Groups Name          Std.Dev.
> individ (Intercept)  53.003
> year2   (Intercept)   1.206
> Number of obs: 210, groups:  individ, 210; year2, 3
> Fixed Effects:
> (Intercept)          sizea2
> 12.83279            0.03985
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Juvenile size model:
> Linear mixed model fit by REML ['lmerMod']
> Formula: sizea3 ~ sizea2 + (1 | year2)
> Data: subdata
> REML criterion at convergence: 1243.682
> Random effects:
> Groups Name          Std.Dev.
> year2   (Intercept)  1.955
> Residual                    5.995
> Number of obs: 193, groups:  year2, 3
> Fixed Effects:
> (Intercept)          sizea2
> 3.0848              0.8465
>
>
> Juvenile secondary size model:
> [1] 1
>
>
> Juvenile tertiary size model:
> [1] 1
>
>
> Juvenile reproduction model:
> [1] 1
>
>
> Juvenile maturity model:

```

```

> [1] 1
>
>
>
>
>
> Number of models in survival table: 2
>
> Number of models in observation table: 2
>
> Number of models in size table: 2
>
> Number of models in secondary size table: 1
>
> Number of models in tertiary size table: 1
>
> Number of models in reproduction status table: 1
>
> Number of models in fecundity table: 4
>
> Number of models in juvenile survival table: 2
>
> Number of models in juvenile observation table: 2
>
> Number of models in juvenile size table: 2
>
> Number of models in juvenile secondary size table: 1
>
> Number of models in juvenile tertiary size table: 1
>
> Number of models in juvenile reproduction table: 1
>
> Number of models in juvenile maturity table: 1
>
>
>
>
> General model parameter names (column 1), and
> specific names used in these models (column 2):
>
>      parameter_names mainparams
> 1          time t      year2
> 2      individual    individ
> 3          patch      patch
> 4      alive in time t+1    surv3
> 5      observed in time t+1    obs3
> 6      sizea in time t+1      size3
> 7      sizeb in time t+1      sizeb3
> 8      sizec in time t+1      sizec3
> 9      reproductive status in time t+1    repst3
> 10     fecundity in time t+1      fec3

```

```

> 11          fecundity in time t          fec2
> 12          sizea in time t            size2
> 13          sizea in time t-1          size1
> 14          sizeb in time t            sizeb2
> 15          sizeb in time t-1          sizeb1
> 16          sizec in time t            sizec2
> 17          sizec in time t-1          sizec1
> 18    reproductive status in time t    repst2
> 19    reproductive status in time t-1  repst1
> 20          maturity status in time t+1 matst3
> 21          maturity status in time t  matst2
> 22          age in time t              age
> 23          density in time t          density
> 24    individual covariate a in time t  indcova2
> 25    individual covariate a in time t-1 indcova1
> 26    individual covariate b in time t  indcovb2
> 27    individual covariate b in time t-1 indcovb1
> 28    individual covariate c in time t  indcovc2
> 29    individual covariate c in time t-1 indcovc1
> 30          stage group in time t      group2
> 31          stage group in time t-1    group1
>
>
>
>
> Quality control:
>
> Survival estimated with 931 individuals and 2246 individual transitions.
> Survival accuracy is 0.977.
> Observation estimated with 858 individuals and 2121 individual transitions.
> Observation accuracy is 0.903.
> Primary size estimated with 845 individuals and 1916 individual transitions.
> Primary size pseudo R-squared is 0.525.
> Secondary size transition not estimated.
> Tertiary size transition not estimated.
> Reproduction probability not estimated.
> Fecundity estimated with 931 individuals and 2246 individual transitions.
> Fecundity pseudo R-squared is 0.236.
> Juvenile survival estimated with 281 individuals and 281 individual transitions.
> Juvenile survival accuracy is 0.747.
> Juvenile observation estimated with 210 individuals and 210 individual transitions.
> Juvenile observation accuracy is 1.
> Juvenile primary size estimated with 193 individuals and 193 individual transitions.
> Juvenile primary size pseudo R-squared is 0.332.
> Juvenile secondary size transition not estimated.
> Juvenile tertiary size transition not estimated.
> Juvenile reproduction probability not estimated.
> Juvenile maturity probability not estimated.

```

7.2.4 Bringing our discretized IPMs to life

The typical IPM is ahistorical and so will utilize only an ahistorical set of vital rate models to populate its matrices. Let's do that and take a look at the result.

```
lathmat2ipm <- flefko2(stageframe = lathframeipm, modelsuite = lathmodels2ipm,
  supplement = lathsupp2, data = lathvertipm, reduce = FALSE)

summary(lathmat2ipm)
>
> This ahistorical lefkoMat object contains 3 matrices.
>
> Each matrix is square with 103 rows and columns, and a total of 10609 elements.
> A total of 26947 survival transitions were estimated, with 8982.333 per matrix.
> A total of 600 fecundity transitions were estimated, with 200 per matrix.
> This lefkoMat object covers 1 population, 1 patch, and 3 time steps.
>
> Vital rate modeling quality control:
>
> Survival estimated with 931 individuals and 2246 individual transitions.
> Observation estimated with 858 individuals and 2121 individual transitions.
> Primary size estimated with 845 individuals and 1916 individual transitions.
> Secondary size transition not estimated.
> Tertiary size transition not estimated.
> Reproduction probability not estimated.
> Fecundity estimated with 931 individuals and 2246 individual transitions.
> Juvenile survival estimated with 281 individuals and 281 individual transitions.
> Juvenile observation estimated with 210 individuals and 210 individual transitions.
> Juvenile primary size estimated with 193 individuals and 193 individual transitions.
> Juvenile secondary size transition not estimated.
> Juvenile tertiary size transition not estimated.
> Juvenile reproduction probability not estimated.
> Juvenile maturity transition probability not estimated.
>
> Survival probability sum check (each matrix represented by column in order):
>      [,1] [,2] [,3]
> Min.  0.000 0.000 0.000
> 1st Qu. 0.994 0.970 0.996
> Median 1.000 1.000 1.000
> Mean  0.961 0.929 0.965
> 3rd Qu. 1.000 1.000 1.000
> Max.  1.000 1.000 1.000
```

The ahistorical IPM is composed of three matrices, covering each of the time steps. These are large matrices - with 103 rows and columns, they include 10,609 elements each. Of these, on average 9182.33 elements in each matrix are non-zero, meaning that these matrices are not only large but also quite dense (86.6% of elements are estimated).

We will now create the historical suite of matrices covering the years of study. These matrices will be extremely large - large enough that some computers might have difficulty with them. If you encounter an error message telling you that you have run out of memory, then please try this on a more powerful computer :) .

```

lathmat3ipm <- flefko3(stageframe = lathframeipm, modelsuite = lathmodels3ipm,
  supplement = lathsupp3, data = lathvertipm, reduce = FALSE)

summary(lathmat3ipm)
>
> This historical lefkoMat object contains 3 matrices.
>
> Each matrix is square with 10609 rows and columns, and a total of 112550881 elements.
> A total of 2684709 survival transitions were estimated, with 894903 per matrix.
> A total of 60600 fecundity transitions were estimated, with 20200 per matrix.
> This lefkoMat object covers 1 population, 1 patch, and 3 time steps.
>
> Vital rate modeling quality control:
>
> Survival estimated with 931 individuals and 2246 individual transitions.
> Observation estimated with 858 individuals and 2121 individual transitions.
> Primary size estimated with 845 individuals and 1916 individual transitions.
> Secondary size transition not estimated.
> Tertiary size transition not estimated.
> Reproduction probability not estimated.
> Fecundity estimated with 931 individuals and 2246 individual transitions.
> Juvenile survival estimated with 281 individuals and 281 individual transitions.
> Juvenile observation estimated with 210 individuals and 210 individual transitions.
> Juvenile primary size estimated with 193 individuals and 193 individual transitions.
> Juvenile secondary size transition not estimated.
> Juvenile tertiary size transition not estimated.
> Juvenile reproduction probability not estimated.
> Juvenile maturity transition probability not estimated.
>
> Survival probability sum check (each matrix represented by column in order):
>      [,1] [,2] [,3]
> Min.  0.000 0.000 0.000
> 1st Qu. 0.993 0.986 0.992
> Median 0.998 0.998 0.998
> Mean  0.958 0.947 0.957
> 3rd Qu. 0.999 0.999 0.999
> Max.  1.000 1.000 1.000

```

These are giant matrices. With 10,609 rows and columns, there are a total of 112,550,881 elements per matrix. But they are also amazingly sparse - with 915,103 elements estimated, only 0.8% of elements per matrix are non-zero. The survival probability sums all look good, so we appear to have no problems with overly large given and proxy survival transitions provided through our supplemental tables.

At this stage, we have created our IPMs. Congratulations! We can also create arithmetic mean matrix versions of each, as below.

```

lath2ipmmean <- lmean(lathmat2ipm)
lath3ipmmean <- lmean(lathmat3ipm)

summary(lath2ipmmean)
>
> This ahistorical lefkoMat object contains 1 matrix.

```



```

>
> Each matrix is square with 103 rows and columns, and a total of 10609 elements.
> A total of 9111 survival transitions were estimated, with 9111 per matrix.
> A total of 200 fecundity transitions were estimated, with 200 per matrix.
> This lefkoMat object covers 1 population, 1 patch, and 0 time steps.
>
> Vital rate modeling quality control:
>
> Survival estimated with 931 individuals and 2246 individual transitions.
> Observation estimated with 858 individuals and 2121 individual transitions.
> Primary size estimated with 845 individuals and 1916 individual transitions.
> Secondary size transition not estimated.
> Tertiary size transition not estimated.
> Reproduction probability not estimated.
> Fecundity estimated with 931 individuals and 2246 individual transitions.
> Juvenile survival estimated with 281 individuals and 281 individual transitions.
> Juvenile observation estimated with 210 individuals and 210 individual transitions.
> Juvenile primary size estimated with 193 individuals and 193 individual transitions.
> Juvenile secondary size transition not estimated.
> Juvenile tertiary size transition not estimated.
> Juvenile reproduction probability not estimated.
> Juvenile maturity transition probability not estimated.
>
> Survival probability sum check (each matrix represented by column in order):
>      [,1]
> Min.    0.000
> 1st Qu. 0.987
> Median  1.000
> Mean    0.952
> 3rd Qu. 1.000
> Max.    1.000
summary(lath3ipmmean)
>
> This historical lefkoMat object contains 1 matrix.
>
> Each matrix is square with 10609 rows and columns, and a total of 112550881 elements.
> A total of 920292 survival transitions were estimated, with 920292 per matrix.
> A total of 20200 fecundity transitions were estimated, with 20200 per matrix.
> This lefkoMat object covers 1 population, 1 patch, and 0 time steps.
>
> Vital rate modeling quality control:
>
> Survival estimated with 931 individuals and 2246 individual transitions.
> Observation estimated with 858 individuals and 2121 individual transitions.
> Primary size estimated with 845 individuals and 1916 individual transitions.
> Secondary size transition not estimated.
> Tertiary size transition not estimated.
> Reproduction probability not estimated.
> Fecundity estimated with 931 individuals and 2246 individual transitions.
> Juvenile survival estimated with 281 individuals and 281 individual transitions.
> Juvenile observation estimated with 210 individuals and 210 individual transitions.

```

```

> Juvenile primary size estimated with 193 individuals and 193 individual transitions.
> Juvenile secondary size transition not estimated.
> Juvenile tertiary size transition not estimated.
> Juvenile reproduction probability not estimated.
> Juvenile maturity transition probability not estimated.
>
> Survival probability sum check (each matrix represented by column in order):
>      [,1]
> Min.   0.000
> 1st Qu. 0.989
> Median 0.998
> Mean   0.954
> 3rd Qu. 0.999
> Max.   1.000

```

7.3 Quality control

IPMs are difficult to inspect because of their size. Package `lefk3` includes a number of ways to assess the overall quality of an IPM. Here, we will cover three main methods, each covering a different aspect of the process. First, we may look at quality control information about our vital rate models. Let's look at a summary of the ahistorical vital rate models.

```

summary(lathmodels2ipm)
> This LefkoMod object includes 7 linear models.
> Best-fit model criterion used: aicc&k
>
>
>
> Survival model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: alive3 ~ (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC    logLik  deviance  df.resid
> 709.9383 727.0890 -351.9691  703.9383    2243
> Random effects:
> Groups Name      Std.Dev.
> individ (Intercept) 14.6
> year2 (Intercept) 0.0
> Number of obs: 2246, groups: individ, 931; year2, 3
> Fixed Effects:
> (Intercept)
>      10.03
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
>
> Observation model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]

```

```

> Family: binomial ( logit )
> Formula: obsstatus3 ~ (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC    logLik deviance df.resid
> 1337.3193 1354.2982 -665.6596 1331.3193     2118
> Random effects:
> Groups Name      Std.Dev.
> individ (Intercept) 1.216
> year2   (Intercept) 0.000
> Number of obs: 2121, groups:  individ, 858; year2, 3
> Fixed Effects:
> (Intercept)
>      2.788
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Size model:
> Linear mixed model fit by REML ['lmerMod']
> Formula: sizea3 ~ sizea2 + (1 | year2) + (1 | individ)
> Data: subdata
> REML criterion at convergence: 29294.15
> Random effects:
> Groups Name      Std.Dev.
> individ (Intercept) 0.0
> year2   (Intercept) 210.9
> Residual                    504.6
> Number of obs: 1916, groups:  individ, 845; year2, 3
> Fixed Effects:
> (Intercept)      sizea2
> 164.0695      0.6211
> optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Secondary size model:
> [1] 1
>
>
> Tertiary size model:
> [1] 1
>
>
> Reproductive status model:
> [1] 1
>
>
> Fecundity model:

```

```

> Formula:          fecae2 ~ sizeae2 + (1 | year2) + (1 | individ)
> Zero inflation:   ~sizeae2 + (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC    logLik  df.resid
> 2905.970 2957.422 -1443.985    2237
> Random-effects (co)variances:
>
> Conditional model:
> Groups Name      Std.Dev.
> year2  (Intercept) 0.1432
> individ (Intercept) 0.3305
>
> Zero-inflation model:
> Groups Name      Std.Dev.
> year2  (Intercept) 0.3898
> individ (Intercept) 1.0047
>
> Number of obs: 2246 / Conditional model: year2, 3; individ, 931 / Zero-inflation model: year2, 3; i
>
> Dispersion parameter for nbinom2 family (): 2.17
>
> Fixed Effects:
>
> Conditional model:
> (Intercept)      sizeae2
> 1.736911      0.000286
>
> Zero-inflation model:
> (Intercept)      sizeae2
> 4.014336      -0.001969
>
>
> Juvenile survival model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: alive3 ~ (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC    logLik  deviance  df.resid
> 323.6696 334.5847 -158.8348 317.6696    278
> Random effects:
> Groups Name      Std.Dev.
> individ (Intercept) 0.0003658
> year2  (Intercept) 0.0000000
> Number of obs: 281, groups:  individ, 281; year2, 3
> Fixed Effects:
> (Intercept)
> 1.084
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>

```

```
>
> Juvenile observation model:
> Generalized linear mixed model fit by maximum likelihood (Laplace
> Approximation) [glmerMod]
> Family: binomial ( logit )
> Formula: obsstatus3 ~ sizea2 + (1 | year2) + (1 | individ)
> Data: subdata
>      AIC      BIC  logLik deviance df.resid
> 61.3733 74.7617 -26.6867  53.3733     206
> Random effects:
> Groups Name          Std.Dev.
> individ (Intercept) 53.003
> year2   (Intercept)  1.206
> Number of obs: 210, groups:  individ, 210; year2, 3
> Fixed Effects:
> (Intercept)          sizea2
> 12.83279            0.03985
> optimizer (Nelder_Mead) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warnings
>
>
> Juvenile size model:
> Linear mixed model fit by REML ['lmerMod']
> Formula: sizea3 ~ sizea2 + (1 | year2)
> Data: subdata
> REML criterion at convergence: 1243.682
> Random effects:
> Groups Name          Std.Dev.
> year2   (Intercept)  1.955
> Residual                    5.995
> Number of obs: 193, groups:  year2, 3
> Fixed Effects:
> (Intercept)          sizea2
> 3.0848              0.8465
>
>
> Juvenile secondary size model:
> [1] 1
>
>
> Juvenile tertiary size model:
> [1] 1
>
>
> Juvenile reproduction model:
> [1] 1
>
>
```

```

>
> Juvenile maturity model:
> [1] 1
>
>
>
>
> Number of models in survival table: 2
>
> Number of models in observation table: 2
>
> Number of models in size table: 2
>
> Number of models in secondary size table: 1
>
> Number of models in tertiary size table: 1
>
> Number of models in reproduction status table: 1
>
> Number of models in fecundity table: 4
>
> Number of models in juvenile survival table: 2
>
> Number of models in juvenile observation table: 2
>
> Number of models in juvenile size table: 2
>
> Number of models in juvenile secondary size table: 1
>
> Number of models in juvenile tertiary size table: 1
>
> Number of models in juvenile reproduction table: 1
>
> Number of models in juvenile maturity table: 1
>
>
>
>
> General model parameter names (column 1), and
> specific names used in these models (column 2):
>           parameter_names mainparams
> 1           time t         year2
> 2         individual      individ
> 3           patch         patch
> 4         alive in time t+1   surv3
> 5     observed in time t+1     obs3
> 6         sizea in time t+1   size3
> 7         sizeb in time t+1   sizeb3
> 8         sizec in time t+1   sizec3

```

```

> 9      reproductive status in time t+1      repst3
> 10      fecundity in time t+1      fec3
> 11      fecundity in time t      fec2
> 12      sizea in time t      size2
> 13      sizea in time t-1      size1
> 14      sizeb in time t      sizeb2
> 15      sizeb in time t-1      sizeb1
> 16      sizec in time t      sizec2
> 17      sizec in time t-1      sizec1
> 18      reproductive status in time t      repst2
> 19      reproductive status in time t-1      repst1
> 20      maturity status in time t+1      matst3
> 21      maturity status in time t      matst2
> 22      age in time t      age
> 23      density in time t      density
> 24      individual covariate a in time t      indcova2
> 25      individual covariate a in time t-1      indcova1
> 26      individual covariate b in time t      indcova2
> 27      individual covariate b in time t-1      indcova1
> 28      individual covariate c in time t      indcova2
> 29      individual covariate c in time t-1      indcova1
> 30      stage group in time t      group2
> 31      stage group in time t-1      group1
>
>
>
>
> Quality control:
>
> Survival estimated with 931 individuals and 2246 individual transitions.
> Survival accuracy is 0.977.
> Observation estimated with 858 individuals and 2121 individual transitions.
> Observation accuracy is 0.903.
> Primary size estimated with 845 individuals and 1916 individual transitions.
> Primary size pseudo R-squared is 0.525.
> Secondary size transition not estimated.
> Tertiary size transition not estimated.
> Reproduction probability not estimated.
> Fecundity estimated with 931 individuals and 2246 individual transitions.
> Fecundity pseudo R-squared is 0.236.
> Juvenile survival estimated with 281 individuals and 281 individual transitions.
> Juvenile survival accuracy is 0.747.
> Juvenile observation estimated with 210 individuals and 210 individual transitions.
> Juvenile observation accuracy is 1.
> Juvenile primary size estimated with 193 individuals and 193 individual transitions.
> Juvenile primary size pseudo R-squared is 0.332.
> Juvenile secondary size transition not estimated.
> Juvenile tertiary size transition not estimated.
> Juvenile reproduction probability not estimated.
> Juvenile maturity probability not estimated.

```

At the very bottom of our output is a section labelled `Quality control`. We see, first of all, a statement of which of our fourteen possible vital rate models were estimated. For each estimated model, we see the number of individuals and actual transitions used to estimate the respective model. In general, the higher the number of individuals and transitions used to estimate the model, the better the quality of the model and the higher the statistical power. The former number, the number of individuals, particularly gives us a sense of the overall level of pseudoreplication that might be inherent in our analysis, since transitions from the same individual are obviously related and not statistically independent of one another.

Our output also includes information on the accuracy of binomial models and pseudo- R^2 of size and fecundity models. Accuracy is calculated as the proportion of predicted responses from a binomial model equal to the actual responses given each data point. Pseudo- R^2 is calculated differently depending on the distribution and modeling form used. Accuracy and pseudo- R^2 both vary from 0.0 to 1.0, and the higher the number the better the quality of the model. What is a “good” quality of model is difficult to say, but prediction should probably not be attempted with vital rate models under 90% accuracy or pseudo- R^2 . Naturally, such values may be difficult to achieve in most analyses.

The next method of assessing quality control focuses on the IPMs, themselves. Let’s take a look at a summary of the ahistorical IPM.

```
summary(lathmat2ipm)
>
> This ahistorical lefkoMat object contains 3 matrices.
>
> Each matrix is square with 103 rows and columns, and a total of 10609 elements.
> A total of 26947 survival transitions were estimated, with 8982.333 per matrix.
> A total of 600 fecundity transitions were estimated, with 200 per matrix.
> This lefkoMat object covers 1 population, 1 patch, and 3 time steps.
>
> Vital rate modeling quality control:
>
> Survival estimated with 931 individuals and 2246 individual transitions.
> Observation estimated with 858 individuals and 2121 individual transitions.
> Primary size estimated with 845 individuals and 1916 individual transitions.
> Secondary size transition not estimated.
> Tertiary size transition not estimated.
> Reproduction probability not estimated.
> Fecundity estimated with 931 individuals and 2246 individual transitions.
> Juvenile survival estimated with 281 individuals and 281 individual transitions.
> Juvenile observation estimated with 210 individuals and 210 individual transitions.
> Juvenile primary size estimated with 193 individuals and 193 individual transitions.
> Juvenile secondary size transition not estimated.
> Juvenile tertiary size transition not estimated.
> Juvenile reproduction probability not estimated.
> Juvenile maturity transition probability not estimated.
>
> Survival probability sum check (each matrix represented by column in order):
>      [,1] [,2] [,3]
> Min.   0.000 0.000 0.000
> 1st Qu. 0.994 0.970 0.996
> Median 1.000 1.000 1.000
> Mean   0.961 0.929 0.965
> 3rd Qu. 1.000 1.000 1.000
> Max.   1.000 1.000 1.000
```


Some of the output should be familiar, particularly the output related to the vital rate models. The key output for us to look at here is at the bottom, under **Survival probability sum check**. The columns in our U matrices should always sum to values between 0 and 1. The reason is that the sum of each column should equal the probability of survival from whatever stage the column is associated with in time t to time $t+1$, regardless of what stage the organism is in in the latter time. In at least three circumstances, these sums may be greater than 1.0, and the user would need to correct their IPMs in these cases to prevent odd analytical results and erroneous inferences. The first circumstance is through the incorporation via supplement tables of fixed transition probabilities or proxy values that are too high. Fixing an IPM in this case would mean reducing these fixed or proxy values in the supplement table. The second circumstance is through the use of the midpoint method in size transition probability estimation. The correct way to fix this is to use the cumulative density function (CDF) method instead of the midpoint method. Fortunately, `lefko3` uses the CDF method by default. The third circumstance is through the incorporation of sizes not observed, or representing strong outlier sizes. In these cases, it is possible for at least some of the resulting probabilities to be estimated at unnatural levels. In this third case, the way to correct the problem is to remove the outlier size from the classification in the stageframe.

Finally, there is at least one more method that we can use to assess the overall quality of an IPM. That method is to assess the overall structure of the IPM. The best way to do this is to inspect the elements themselves, perhaps by opening the IPM matrix in R Studio, or exporting it to Microsoft Excel or another spreadsheet program for assessment. A more visual approach assessing just the structure itself is to use the `image3()` function, which provides users with a means of assessing whether the overall structure of the model “looks right”. For example, here we look at an image of the 1st matrix in our ahistorical IPM (figure 7.5).

```
image3(lathmat2ipm, used = 1)
```

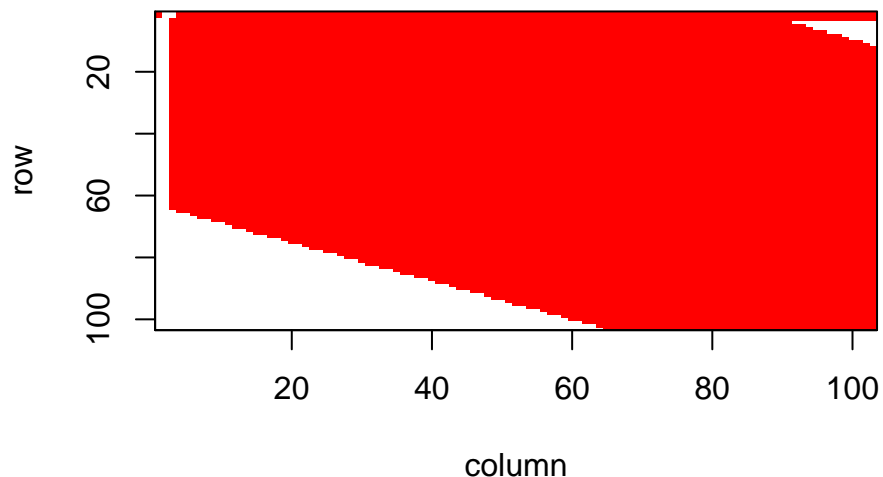


Figure 7.5: Matrix image of the first ahistorical IPM. Red area corresponds to non-zero elements

```
> [[1]]
> NULL
```

We can also focus in on just the survival or fecundity transitions, as below. Note that in an ahistorical IPM, we expect all fecundity values to be located toward the top of the matrix (figure 7.6).

```
image3(lathmat2ipm, used = 1, type = "F")
```

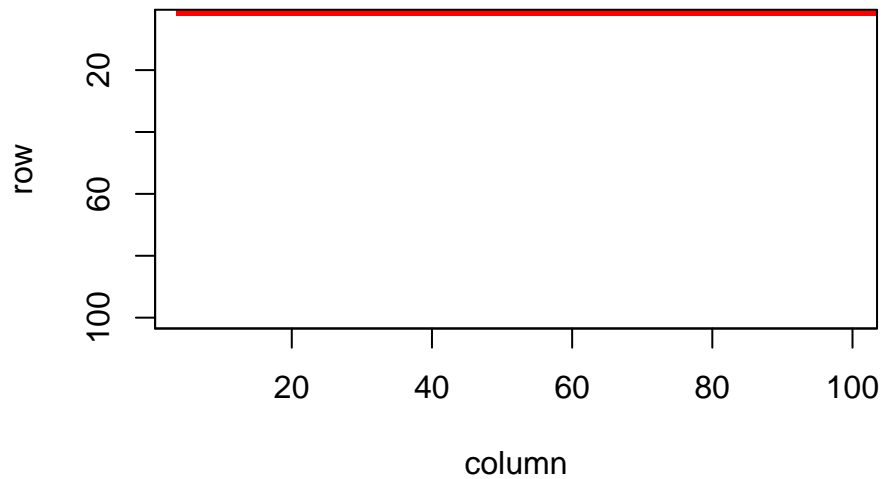


Figure 7.6: Matrix image of the first fecundity matrix in our ahistorical IPM. Red area corresponds to non-zero elements

```
> [[1]]
> NULL
```

Other approaches to quality control are provided for other aspects of matrix analysis in `lefko3`.

At this point, we may move ahead and analyze the IPMs in the same ways that we might analyze other kinds of MPMs.

7.4 Points to remember

1. IPMs assume a continuous distribution for size and integrate vital rates across this size metric typically assuming a Gaussian size distribution. While theoretically different from MPMs, ultimately IPMs are discretized in ways that allow them to be created and analyzed exactly the same way as function-based MPMs.
2. Package `lefko3` allows the development of even extremely large numbers of size bins across the range of a continuous size metric, using the "ipm" shorthand in the `sf_create()` function. With only a few lines, a stageframe with hundreds of discretized size bins can be created and used to create an IPM for analysis.
3. Quality control tools include the linear model accuracy and pseudo- R^2 output from `modelsearch()`, the output from `summary()` calls for IPMs created with `flefko2()` or `flefko3()`, and visualization with `image3()`.