

Brazil data

Vic Quennessen

2025-11-13

1. are all males represented in all clutches of females they mate with in a season?

```
clean_data %>%  
  group_by(Season, Female) %>%  
  filter(length(unique(Nest)) > 1)
```

```
## # A tibble: 0 x 9  
## # Groups:   Season, Female [0]  
## # i 9 variables: mom_nest_combo <chr>, dad <chr>, hatchling_count <dbl>,  
## #   NestTotalHatchCount <dbl>, proptadcontribute <dbl>,  
## #   LowNestSampleSize <chr>, Season <chr>, Female <chr>, Nest <chr>
```

Not applicable, no more than one clutch is represented per female per season in the dataset.

I will therefore assume that all males are represented in all clutches by the same female throughout the same season.

2. how many males do females mate with in one season?

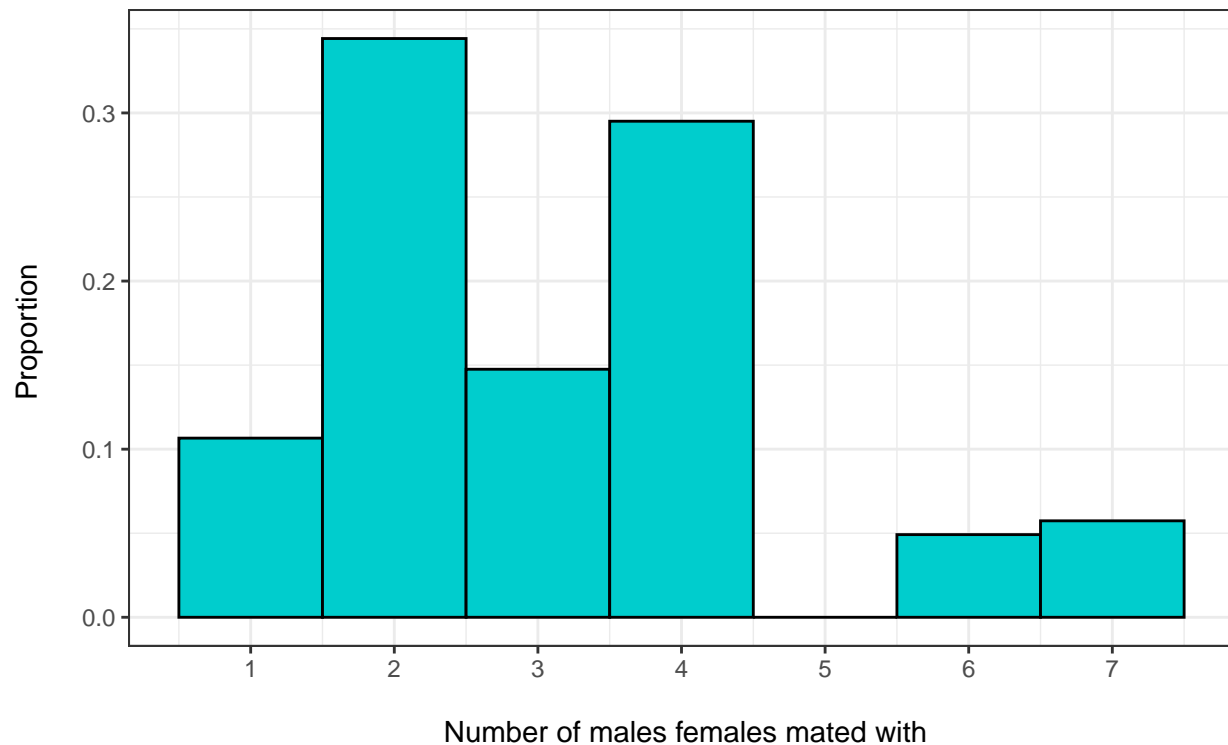
In order to more accurately model evolution with maternal and paternal genotypes, I will build a ‘breeding pool’ of males, where each male will be represented X times, where X is the average number of females that males can mate with, rounded to the nearest integer, dependent upon the mating function. Then, for each female, I need to be able to determine how many males she will mate with to draw that many unique males from the breeding pool.

Number of males	1	2	3	4	5	6	7
Number	13	42	18	36	0	6	7
Proportion	0.11	0.34	0.15	0.30	0	0.05	0.06

Number of males that females mated with in raw data:

Min	1st Q	Median	Mean	3rd Q	Max	Variance	SD	Skew	Kurtosis
1.000	2.000	3.000	3.115	4.000	7.000	2.433	1.560	0.861	0.265

Summary statistics of number of males females mated with: Density histogram



why fit a model

we can't just use the proportions of females that mated with 1 - 7 males since sample sizes are small, and no recorded female mated with 5 males.

Distribution	Log likelihood	AIC	BIC
discrete gamma	-215.0036	434.0071	439.6152
poisson	-223.1883	448.3767	451.1807
negative binomial	-223.1884	450.3767	455.9848
geometric	-278.3816	558.7632	561.5673

discrete gamma distribution had the best fit:

```
# model with fitted discrete gamma distribution
fit_dgamma <- fitdist(males_per_female$nMales, 'gamma', discrete = TRUE)

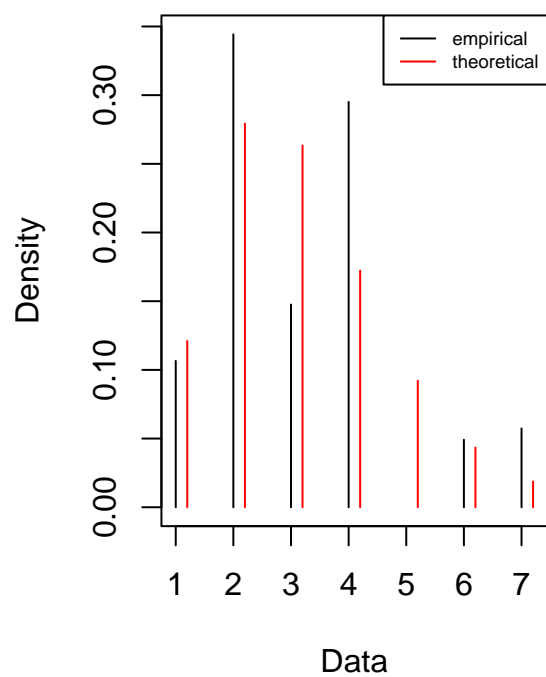
# estimates for shape and rate parameters for discrete gamma distribution
summary(fit_dgamma)$estimate
```

model fit:

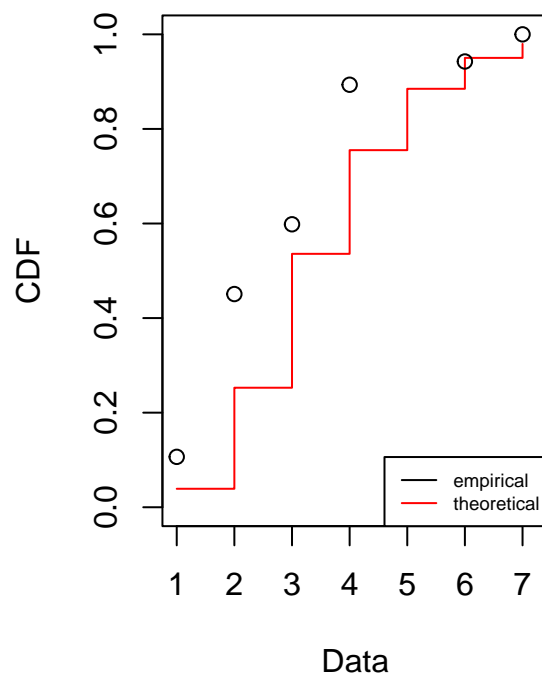
```
##      shape      rate
## 4.109537 1.319252
```

```
# plot data (black) vs. fitted discrete gamma distribution (red)
plot(fit_dgamma)
```

Emp. and theo. distr.



Emp. and theo. CDFs



given this discrete model... what are the weights for females mating with 1 X males, where X is the maximum number of males we will allow (currently set to 10):

```
# max number of males females can mate with (7 in data)
max_males <- 10

# weighted probabilities for females mating with 1 - 7 males
weightedP <- ddgamma(1:max_males,
                     shape = fit_dgamma$estimate['shape'],
                     rate = fit_dgamma$estimate['rate'])

raw_weights <- round(weightedP, 3)
# 0.214 0.283 0.219 0.130 0.065 0.030 0.012

total_sum <- sum(raw_weights)
# 0.9531114

# normalize weights so that they add up to 1
# this assumes we can't have females mating with more than max_males males)
# to adjust this assumption, adjust max_males values
weights_to_use <- round(weightedP / sum(weightedP), 3)
probability_of_mating_with_n_Males <- data.frame(number_of_males = 1:max_males,
                                                  probability = weights_to_use)

probability_of_mating_with_n_Males
```

##	number_of_males	probability
## 1	1	0.222
## 2	2	0.295
## 3	3	0.228
## 4	4	0.135
## 5	5	0.068
## 6	6	0.031
## 7	7	0.013
## 8	8	0.005
## 9	9	0.002
## 10	10	0.001

3. how do males fertilize nests?

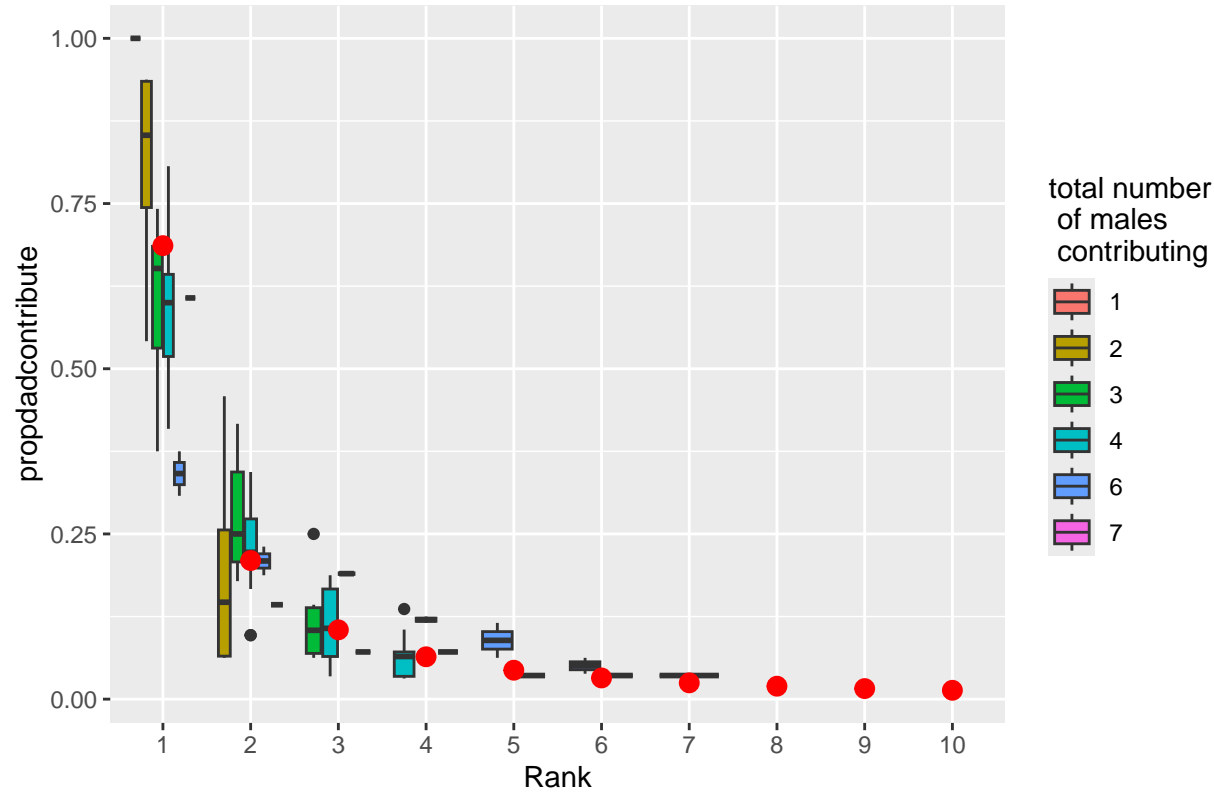
decaying power law fits very well: $y(x) = a \cdot p^{-x}$ Where the intercept = -0.37648 and the log(Rank) = -1.71013:

```
# turtles work like gravitational waves - jerry sun
# power law
fit_power <- lm(log(propdadcontribute) ~ log(Rank),
               data = contributions)
summary(fit_power)
```

```
##
## Call:
## lm(formula = log(propdadcontribute) ~ log(Rank), data = contributions)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21074 -0.23495  0.08339  0.36808  0.96934
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.37648    0.05858  -6.427 1.78e-09 ***
## log(Rank)   -1.71013    0.07091 -24.118 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4777 on 144 degrees of freedom
## (3 observations deleted due to missingness)
## Multiple R-squared:  0.8016, Adjusted R-squared:  0.8002
## F-statistic: 581.7 on 1 and 144 DF, p-value: < 2.2e-16
```

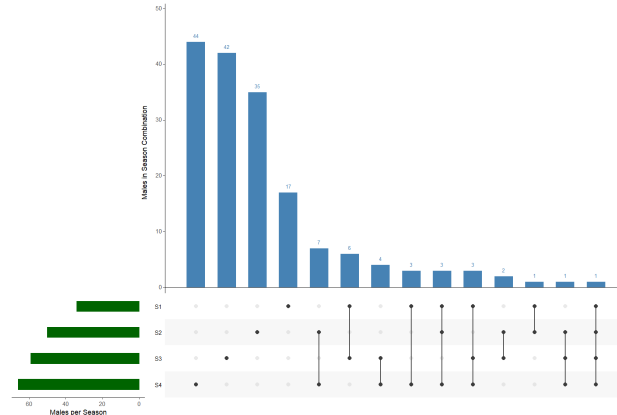
Model fit is plotted in red points:

Fertilization contributions by Rank (Rank 1 = dominant father)



4. how often do males remigrate

male remigration interval used in Chapter 2: 1.47 (years) for green turtle males at Atol das Rocas (Grossman et al., 2019 <https://www.int-res.com/abstracts/meps/v609/p197-207/>)



FdN genetic data

Original data, obtained from Dr. Lisa Komoroske (lkomoroske@umass.edu) on Thursday October 9, 2025. Note: "... Estefany made this upsetR graph for Mariana for an upcoming talk, helps visualize the observed remigration interval a bit for the males (again, keep in mind the difference in sampling effort in S3 and S4, these will need to be normalized to the number of samples run eventually).

fit a binomial model

marginalized likelihoods over number of 0 and 1 successes Given a binomial model, or any resident male green turtle, the probability of seeing it $k = 1, 2, 3,$ or 4 times across n years is equal to:

$$L(k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

We know that $n = 4$, since our data were collected over 4 years, giving 4 chances to observe each male. 66 males were observed in season 4, meaning they could have been observed for all previous seasons. We can assume that those males that were observed more than once were residents. For the remaining 44 males, some unknown number of them were residents and others were transient males. For RM1 resident males observed once, there were $RM1 + 22$ total resident males observed in our dataset.

Given probability p , the likelihoods of observing RM1 males once, 14 males twice, 7 males 3 times, and 1 male 4 times is the probability of seeing a resident male 1, 2, 3, or 4 times in any given trial of size 4 to the power of the number of times it was observed:

And the total likelihood of seeing all of these is the product of each probability times the number of potential combinations of the order in which they were observed:

The conditional likelihood is then the probability of observing RM1 males once out of all males observed ($RM1 + 22$) given probability p :

The marginalized likelihood for each value of N is then the likelihood times the conditional likelihood, and the total likelihood for each probability P is equal to the sum of all the marginalized likelihoods across all potential values of N .

```
## # A tibble: 2 x 3
## # Groups:   Analysis [2]
```

```
## Analysis      P Likelihood
## <chr>         <dbl>         <dbl>
## 1 S4 males    0.55 0.000288
## 2 all males   0.52 0.00000378
```

Analysis	Log likelihood	Probability
All males	0.00000378	0.52
Season 4 males only	0.000288	0.55

If we compare the results from analysing only those resident males observed in Season 4 (as opposed to all resident males), the results do not substantially change. The maximum likelihood estimator for P drops from 0.55 to 0.52, but the overall likelihoods are much smaller.

```
S4P <- results %>%
  filter(Analysis == 'S4 males') %>%
  filter(Likelihood == max(Likelihood)) %>%
  pull(P)

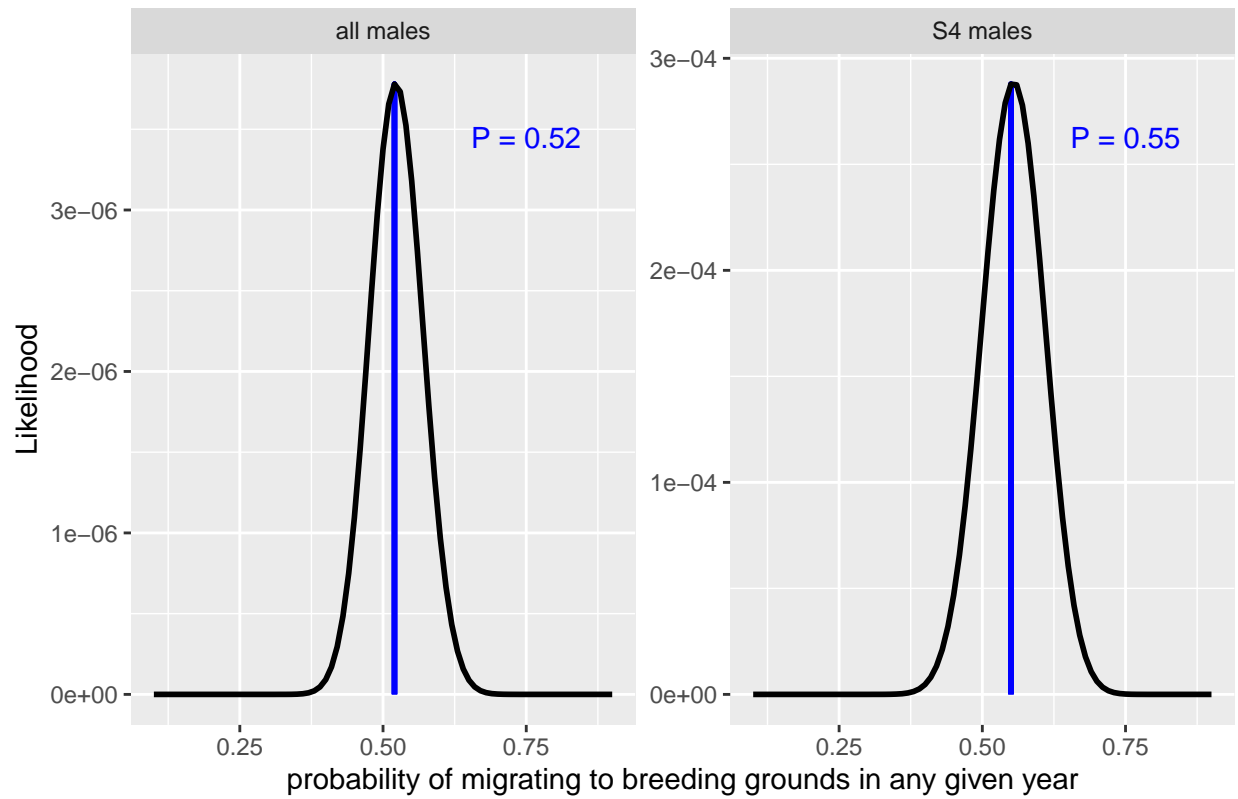
allMP <- results %>%
  filter(Analysis == 'all males') %>%
  filter(Likelihood == max(Likelihood)) %>%
  pull(P)

to_plot <- results %>%
  mutate(maxL = ifelse(Analysis == 'S4 males', S4P, allMP))

fig_labels <- data.frame(Analysis = c('all males', 'S4 males'),
  label = c('P = 0.52', 'P = 0.55'))

ggplot(data = to_plot,
  aes(x = P, y = Likelihood)) +
  geom_line(aes(x = maxL, y = Likelihood),
    col = 'blue', lwd = 1) +
  geom_path(lwd = 1) +
  facet_wrap(vars(Analysis),
    scales = 'free') +
  xlab('probability of migrating to breeding grounds in any given year') +
  ylab('Likelihood') +
  ggtitle('Maximum likelihood estimation for FdN male probability of migrating') +
  geom_text(data = fig_labels,
    mapping = aes(x = 0.75, y = Inf, label = label),
    vjust = 4.5,
    col = 'blue')
```

Maximum likelihood estimation for FdN male probability of migrating



The final maximum likelihood estimator for the binomial probability of males mating is therefore $p = 0.55$, with an associated average remigration interval of 1.82 years.