Introduction to Stochasticity and Probability Theory

Quantitative Analysis of Vertebrate Populations

Why Variation Matters

- Traditionally just a random (normal) nuisance
- Transform to linear or use non-parametric but limits inference
- Tells us about the world: counts following negative binomial indicate undescribed environmental variation or aggregate response not accounted for

Types of "Noise" (stochastic variation)

- Measurement error: results in large CI and low statistical power
- Process variation or noise: actual part of the system
- *Environmental Stochasticity*: spatial and temporal variability caused by the environment rather than inherent randomness in the individuals
- *Demographic Stochasticity*: random chance of a particular animal living or dying ("coin toss")

Probability Theory

1. If two events are mutually exclusive then the prob that either occurs (prob of A or B: $Prob(A \bigcup B)$) is sum of the individual probabilities:

$$Prob(male \bigcup female) = Prob(male) + Prob(female)$$

$$P(3 \leq X \leq 5) = P(X = 3) + P(X = 4) + P(X = 5)$$

Probability Theory

2. If two events, A and B are not mutually exclusive - the *joint probability* that they occur together, $Prob(A \bigcap B)$ is greater than zero - and we have to account for double-counting

$$Prob(A \bigcup B) = Prob(A) + Prob(B) - Prob(A \bigcap B)$$

For example Prob(blue or male) = Prob(blue) + Prob(male) - Prob(blue male)

Probability Theory

3. The probabilities of all possible outcomes of an observation or experiment sum to 1.

Prob(male) + Prob(female) = 1

Probability Theory

4. The *conditional probability* of A given B, $Prob(A|B)$, is the probability that A happens if we know or assume B happens

$$Prob(A|B) = \frac{Prob(A \bigcap B)}{Prob(B)}$$

Probability Theory

5. If the conditional probability of A given B equals the unconditional probability of A, then A is independent of B so

$$Prob(A \bigcap B) = Prob(A)Prob(B)$$

We will refer back to these rules to understand probabilities and stochastic processes later.

Types of Distributions

- **Discrete**: the outcomes are a set of integers, usually counts resulting in non-negative integers in ecology
- **Continuous**: all real values or all real non-negative values (e.g. mass, length)

Characteristics of Distributions: Central Moments

First Moment

Mean (expectations)

$$\bar{x} = E[x] = \frac{\sum_{i=1}^{N} x_i}{N}$$

Second Moment

Variances (Expectations of X^2)

$$E[x - x^2] = E[x^2] - (\bar{x})^2$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}$$

Third Moment

Skewness - how asymmetric a distribution is around its mean

$$E[x - x^3] = E[x^3] - (\bar{x})^3$$

Fourth Moment

Kurtosis - how flat or pointy a distribution is

$$E[x - x^4] = E[x^4] - (\bar{x})^4$$

Non-moment Characteristics

- Median: point that divides the area of the probability density in half or that the cumulative distribution function is 0.5. Less responsive to outliers than the mean
- Mode: the most likely value, the maximum of the probability distribution or density function.
- Symmetric distributions mean = median = mode
- Asymmetric: right skewed generally mode < median < mean

Common Distributions in Population Biology

- Binomial (discrete, 0, N)
- Poisson (discrete, 0, ∞)
- Negative Binomial (discrete, 0, ∞)
- Uniform (continuous, 0, 1)
- Normal/Gaussian (continuous, -∞, ∞)
- Exponential (continuous, 0, ∞)
- Lognormal (continuous, 0, ∞)

Common Distributions in Population Biology

Binomial

Fixed number of samples or "trials", each with only 2 possible outcomes (coin flips with biased coins)

- Distribution: $\left(\begin{array}{c} N \\ x \end{array} \right) p^x (1-p)^{N-x}$
- mean: Np
- variance: $Np(1-p)$

```
Make binomial distibutions in R with 10 trials but with p = 0.1, 0.5, and 0.9
```

Poisson

120 • CHAPTER 4

TABLE 4.1
Summary of Probability Distributions

Distribution	Type	Range	Skew	Examples
Binomial	Discrete	$0, N$	Any	Number surviving, number killed
Poisson	Discrete	$0, \infty$	Right	Seeds per quadrat, settlers (variance/mean ≈ 1)
Negative binomial	Discrete	$0, \infty$	Right	Seeds per quadrat, settlers (variance/mean > 1)
Geometric	Discrete	$0, \infty$	Right	Discrete lifetimes
Beta-binomial	Discrete	$0, N$	Any	Similar to binomial
Uniform	Continous	0, 1	None	Cover proportion
Normal	Continuous	$-\infty, \infty$	None	Mass
Gamma	Continuous	$0, \infty$	Right	Survival time, distance to nearest edge
Beta	Continuous	0, 1	Any	Cover proportion
Exponential	Continuous	$0, \infty$	Right	Survival time, distance to nearest edge
Lognormal	Continuous	$0, \infty$	Right	Size, mass (exponential growth)

Figure 1: Bolker 2008. Ecological Models and Data in R