Outline

1. Probability Model
   - Motivation
   - Experiments and events
   - Rules for probability calculations
   - Independence
2. Random Variables
   - Definition and probability distribution
   - Expectation and standard deviation
   - Cumulative distribution function

Parasitic Fish

Experiment from example 9.3 (p. 213): fish are placed in a large tank for a period of time and some are eaten by large birds of prey. The fish are categorized by their level of parasitic infection: uninfected, lightly infected, or highly infected.

It is to the parasites advantage to be in a fish that is eaten: opportunity to infect the bird in the parasites' next stage of life. Different proportions of fish eaten observed by category:

	Uninfected	Lightly Infected	Highly Infected	Total
Eaten	1	10	37	48
Not eaten	49	35	9	93
Total	50	45	46	141

Proportions of eaten fish:

$$1/50 = 0.02,\ 10/45 = 0.222, \text{ and } 37/46 = 0.804.$$

Questions

There are 3 conditional probabilities of interest: the probability of being eaten by a bird given one of the 3 infection level.

How to test if these are the same?

Estimate how different they are?

Real association between infection level and bird predation?

Motivation

To understand the methods for comparing probabilities in different populations, we need to develop notions of:

conditional probability, and

independence

Tools to formalize ideas of what we “expect just by chance”.

Elements in a probability model

Experiment: action/process that generates data. Usually has more than one possible outcomes, is theoretically repeatable.

Elementary outcome: complete description of a single result from the experiment. Can be quite complicated, but cannot be divided further.

Sample space: the entire group of elementary outcomes, $\mathcal{S}$.

Event: a collection of elementary outcomes, i.e. subset of $\mathcal{S}$.

Elements in a probability model

Example: roll a die once and record the result.

"3" is an elementary outcome;

$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$;

$\{2, 4, 6\}$ is the event of getting an even number.

Example: pick a random fish at random from a tank and record its infection status and its eaten/not eaten fate.

"Uninfected not Eaten" (U, nE) is an elementary outcome;

$\mathcal{S} = \{(U, E), (U, nE), (L, E), (L, nE), (H, E), (H, nE)\}$;

$\{(U, E), (L, E), (H, E)\}$ is the event of sampling a fish that will be eaten.

Operations on events

Union “U or V": elementary outcomes in U, in V, or in both. Written as $U \cup V$.
"Uninfected" or "Eaten" $= \{(U,E),(U,nE),(L,E),(H,E)\}$.

Intersection “U and V": elementary outcomes in both U and V. Written as $U \cap V$.
"Uninfected" and "Eaten" $= \{(U,E)\}$.

Complement “not U": all elementary outcomes in $\mathcal{S}$ that are not in U. Written as $\bar{U}$.
if $\mathcal{U} =$ “Uninfected” $= \{(U,E),(U,nE)\}$,
"not $\mathcal{U}$" $= \{(L,E),(L,nE),(H,E),(H,nE)\} =$ "Lightly or Highly infected".

Two events are mutually exclusive if they do not have any elementary outcomes in common.
Events "Uninfected" and "Eaten" are not mutually exclusive.
Other examples?

Probability model

A probability model consists of a probability assignment to each of the events in $\mathcal{S}$.

Basic rules an assignment must follow

(i) For any event U, $0 \leq \mathbb{P}(U) \leq 1$
(ii) $\mathbb{P}(\mathcal{S}) = 1$
(iii) Addition rule: If U and V are mutually exclusive, then

$$\mathbb{P}(U \text{ or } V) = \mathbb{P}(U) + \mathbb{P}(V)$$

Rules consistent with our intuitive notion of chance.

Example

Theoretical fish tank

Assume these elementary outcome probabilities, which do sum up to 1:

$$\begin{array}{ccc} \mathbb{P}\{(U,E)\} = 0.05 & \mathbb{P}\{(L,E)\} = 0.20 & \mathbb{P}\{(H,E)\} = 0.25 \\ \mathbb{P}\{(U,nE)\} = 0.05 & \mathbb{P}\{(L,nE)\} = 0.30 & \mathbb{P}\{(H,nE)\} = 0.15 \end{array}$$

Following the additivity rule:

$$\begin{aligned} \mathbb{P}(\text{Uninfected}) &= \mathbb{P}\{(U,E) \text{ or } (U,nE)\} \\ &= \mathbb{P}\{(U,E)\} + \mathbb{P}\{(U,nE)\} = 0.05 + 0.05 = 0.10 \\ \mathbb{P}(\text{Eaten}) &= \mathbb{P}\{(U,E) \text{ or } (L,E) \text{ or } (H,E)\} \\ &= \mathbb{P}\{(U,E)\} + \mathbb{P}\{(L,E)\} + \mathbb{P}\{(H,E)\} = 0.50 \end{aligned}$$

Be careful with the additivity rule:

$$\mathbb{P}(\text{Uninfected or Eaten}) = 0.55 \neq \mathbb{P}(\text{Uninfected}) + \mathbb{P}(\text{Eaten}) =$$

Derived rules

What if we cannot apply (iii) to compute $\mathbb{P}(A \text{ or } B)$ because A and B are not mutually exclusive? From (i)–(iii):

(iv) For any two events U, V,

$$\mathbb{P}(U \text{ or } V) = \mathbb{P}(U) + \mathbb{P}(V) - \mathbb{P}(U \text{ and } V)$$

(iv) is consistent with (iii). Now

$$\begin{aligned}\mathbb{P}(\text{Uninfected or Eaten}) &= \\ \mathbb{P}(\text{Uninfected}) + \mathbb{P}(\text{Eaten}) &- \mathbb{P}(\text{Uninfected and Eaten}) \\ &= 0.10 + 0.50 - 0.05 = 0.55\end{aligned}$$

Derived rules

U and "not U" are always mutually exclusive. By (iii),

$$\mathbb{P}(U) + \mathbb{P}(\text{not } U) = \mathbb{P}(\mathcal{S}) = 1$$

Thus the derived rule:

(v) For any event U, $\qquad \mathbb{P}(\text{not } U) = 1 - \mathbb{P}(U)$

Ex:

$$\begin{aligned}\mathbb{P}(\text{Lightly or Highly infected}) &= \mathbb{P}(\text{not Uninfected}) \\ &= 1 - \mathbb{P}(\text{Uninfected}) = 0.95\end{aligned}$$

Conditional probability

Additional information can alter the probability of an event.
Ex: knowing that the fish is highly infected might alter the probability of the fish being eaten.

The conditional probability of an event U given V, $\mathbb{P}(U|V)$, is the probability of U given (or knowing) that V has occurred:

$$\mathbb{P}(U|V) = \frac{\mathbb{P}(U \text{ and } V)}{\mathbb{P}(V)},$$

provided that $\mathbb{P}(V) \neq 0$.

$$\mathbb{P}\{\text{Eaten} \mid \text{Highly infected}\} = \frac{\mathbb{P}\{(E,H)\}}{\mathbb{P}\{\text{Highly}\}} = \frac{0.25}{0.25 + 0.15}$$

$$= 0.625 = 62.5\%$$

Independence

Two events U and V are independent if information about one does not affect the other, that is,

$\mathbb{P}(U|V) = \mathbb{P}(U)$, or

$\mathbb{P}(V|U) = \mathbb{P}(V)$, or equivalently,

$\mathbb{P}(U \text{ and } V) = \mathbb{P}(U)\mathbb{P}(V)$.

This gives us the

Multiplication rule

If U and V are independent, then

$$\mathbb{P}(U \text{ and } V) = \mathbb{P}(U) \times \mathbb{P}(V)$$

Independence

Are "Highly infected" and "Eaten" independent in our theoretical fish tank?
From earlier: $\mathbb{P}\{E|H\} = 0.625$ while $\mathbb{P}\{E\} = 0.50$.
Answer:
Equivalently, look at $\mathbb{P}\{H \text{ and } E\} = 0.25$ and compare to $\mathbb{P}\{H\} * \mathbb{P}\{E\} = (0.25 + 0.15) * 0.50 = 0.20$.

Are "Uninfected" and "Eaten" independent? Answer:

Independence

If we make the assumption that there is **no relationship** = independence between infection status and bird predation, and if we somehow knew that $\mathbb{P}\{\text{Highly infected}\} = 0.40$ and $\mathbb{P}\{\text{Eaten}\} = 0.50$, then we can use the multiplication rule to predict $\mathbb{P}(\text{Highly infected and Eaten}) = 0.40 * 0.50 = 0.20$.

Outline

1 Probability Model
- Motivation
- Experiments and events
- Rules for probability calculations
- Independence

2 Random Variables
- Definition and probability distribution
- Expectation and standard deviation
- Cumulative distribution function

Random Variables - Definition

A random variable (RV) is a variable that depends on the outcome of a chance situation.

A RV is often denoted by capital letters (e.g. Y).

$$\text{Elementary outcome} \longrightarrow Y \text{ value}$$

Example: $Y = 1$ if eaten, 0 otherwise

Pick 1 fish at random and set $Y = 1$ if the fish is eaten by a bird or $Y = 0$ if the fish is not eaten. 6 elementary outcomes:

elementary outcome	probability	Y value
U,E	0.05	1
U,nE	0.05	0
L,E	0.20	1
L,nE	0.30	0
H,E	0.25	1
H,nE	0.15	0

2 values for Y: 0 and 1.
$\mathbb{P}\{Y = 0\} = 50$ and $\mathbb{P}\{Y = 1\} = 0.50$.

Example: Z = # fish eaten by a bird

Pick 2 fish at random. $6 * 6 = 36$ elementary outcomes and 3 values for Z: 0,1 and 2.

```
elementary outcome Y1 Y2  Z value
for fish1 & fish2
  U,E    U,E        1  1    2
  U,E    U,nE       1  0    1
  U,E    L,E        1  1    2
  U,E    L,nE       1  0    1
  U,E    H,E        1  1    2
  U,E    H,nE       1  0    1
  U,nE   U,E        0  1    1
  U,nE   U,nE       0  0    0
  U,nE   L,E        0  1    1
...
  H,nE   L,nE       0  0    0
  H,nE   H,E        1  1    1
  H,nE   H,nE       0  0    0
```

Discrete RV and probability distribution

RV's can be discrete or continuous variables, but not categorical.

The probability distribution of a **discrete** RV is described by the probability of each possible value of the RV

The probability distribution of a **continuous** RV is described by a density curve.

Example: Y = Number of fish eaten by birds

Suppose we sample 3 fish *independently* of each other, and assume a 60% predation rate.
Use $\times$ = eaten, o = not eaten

$$
\begin{aligned}
\mathbb{P}\{Y = 3\} &= \mathbb{P}(\times \times \times) = \mathbb{P}(\times) * \mathbb{P}(\times) * \mathbb{P}(\times) \\
&= \qquad\qquad\qquad\qquad = .216
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{P}\{Y = 2\} &= \mathbb{P}\{\times \times o \text{ or } \times o \times \text{ or } o \times \times\} \\
&= \mathbb{P}\{\times \times o\} + \mathbb{P}\{\times o \times\} + \mathbb{P}\{o \times \times\} \\
&= \mathbb{P}(\times)\mathbb{P}(\times)\mathbb{P}(o) + \mathbb{P}(\times)\mathbb{P}(o)\mathbb{P}(\times) + \mathbb{P}(o)\mathbb{P}(\times)\mathbb{P}(\times) \\
&= \\
&= 3 * 0.144 = 0.432
\end{aligned}
$$

Similarly $\mathbb{P}\{Y = 1\} = \qquad\qquad = 3 * 0.096 = 0.288$
and $\mathbb{P}\{Y = 0\} = \qquad\qquad = 0.064.$

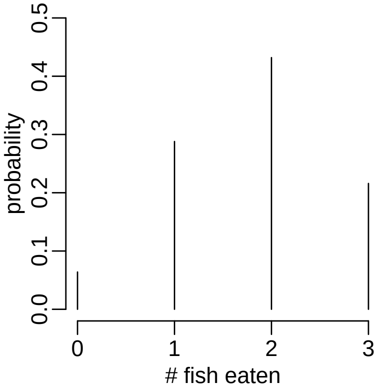
Example: Y = Number of fish eaten by birds

Frequency table for Y:

y	0	1	2	3
p(y)	.064	.288	.432	.216

A **line graph** shows the probability distribution of Y

better than a histogram to explicitely display a **discrete** distribution.



Summary measures

The probability distribution of a discrete RV Y gives complete information about Y and hence complete information about the population.

Helpful to have some numerical summaries such as the center/location or spread/variability of the population (as with sample data).

	population (RV)	sample (observed data)
mean	μ_Y	$\bar{y}$
variance	σ_Y^2	s^2
standard deviation	σ_Y	s

Expectation of a random variable

Ex: pick 3 random fish, independently. Repeat 1000 times.

With a bird predation rate of 60% we got

y	0	1	2	3
p(y)	.064	.288	.432	.216

Roughly, we will find 0 eaten fish 64 times, ...

So we expect to find a total # of destroyed nests of:

$$0 * 64 + \qquad\qquad = 1800$$

and the average # of fish eaten by birds per experiment is:

. This is the expected value of Y.

Expectation of a RV

The expectation of a RV Y is the population mean of the probability distribution of Y. Denoted as $\mathbb{E}(Y)$ or μ_Y. Can be thought of as a typical value.

For a discrete RV Y, it is

Expectation

$$\mathbb{E}(Y) = \sum y \times \mathbb{P}\{Y = y\}$$

summing over all possible values y of the RV Y.

In the fish / bird predation problem:

$$\mathbb{E}(Y) = 0 * 0.064 + 1 * 0.288 + 2 * 0.432 + 3 * 0.216 = 1.8 \text{ fish}$$

Variance and Standard deviation of a RV

The variance of a RV Y – noted $\text{var}(Y)$ or σ_Y^2 – measures the population spread/variability of the distribution of Y. For a discrete RV Y, it is

Variance

$$\text{var}(Y) = \mathbb{E}(Y - \mu_Y)^2 = \sum (y - \mu_Y)^2 \times \mathbb{P}\{Y = y\}$$

summing over all possible values y of the RV Y.
The standard deviation of Y is $\sigma_Y = \sigma = \sqrt{\text{var}(Y)}$.
σ_Y measures the amount Y typically deviates from μ_Y.

Fish eaten by birds, out of 3 sampled fish:

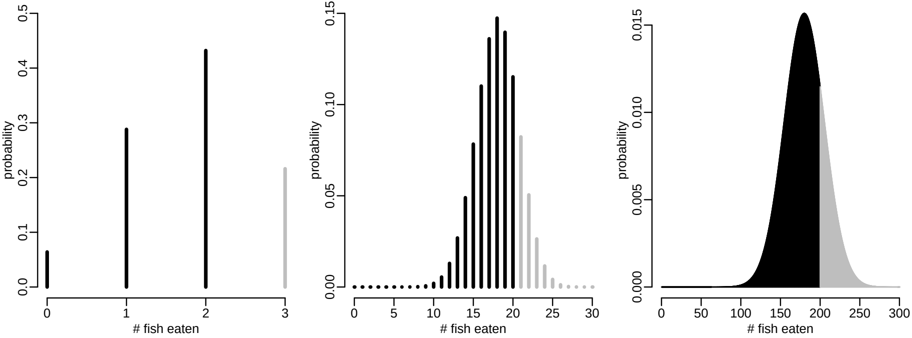
$$\begin{aligned}\text{var}(Y) &= (0 - 1.8)^2 * 0.068 + (1 - 1.8)^2 * 0.288 \\ &\quad + (2 - 1.8)^2 * 0.432 + (3 - 1.8)^2 * 0.216 \\ &= 0.72\end{aligned}$$

and $\sigma_Y = \sqrt{0.72} = 0.848$ fish.

Cumulative distribution function

Answer to "What is the probability that Y is $\leq$ some y value?"

ex: $\mathbb{P}\{Y \leq 2\} = 0.068 + 0.288 + 0.432 = 0.784$



Continuous RV: density curve describes probability distribution, where area = probability. Image: blocks in histogram = ice cubes, ground into fine dust and spread onto a line.