1. Basics

Discrete example

▷ How should we decide whether the coin is fair or .8-heads biased?

▷ A natural idea: flip it, say, 5 times, count the number of heads, and decide on that basis.

▷ What are the probabilities?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>coin is fair</td>
<td>.031</td>
<td>.156</td>
<td>.313</td>
<td>.156</td>
<td>.031</td>
<td></td>
</tr>
<tr>
<td>coin is .8-heads biased</td>
<td>.000</td>
<td>.006</td>
<td>.051</td>
<td>.205</td>
<td>.410</td>
<td>.328</td>
</tr>
</tbody>
</table>

▷ For each outcome, which hypothesis should we reject?

Continuous example

▷ A pressure sensor sends signal 2 if it detects an intruder and 0 if not. But the signal is corrupted in transmission due to random noise, \( N(0, 1) \). How should we decide whether there’s an intruder based on the signal received?

▷ What are the probabilities?

![Figure 2: Densities of outcomes under the two hypotheses. Probabilities are areas under the curves. For example, the probability the received signal is between 0 and 2 is the area under the blue curve between 0 and 2, if there is an intruder, and the area under the orange curve between 0 and 2, if there isn’t.](image)

▷ For each outcome, which hypothesis should we reject?
Definitions

- We aim to decide between two hypotheses, $h$ and $i$, based on the outcome of an experiment.

- A rejection region is a set of outcomes in which we reject $h$ and accept $i$.

- Which rejection region should we use?

  - Good cases and bad cases:

    | reject $h$, accept $i$ | $h$ and $¬i$ | type I error | $¬h$ and $i$ | OK |
    | reject $i$, accept $h$ | OK | type II error |

  - We ignore the case in which both $h$ and $i$ are false.

- The size of a rejection region is the probability of getting an outcome in the region if $h$ is true.

- The power of a rejection region is the probability of getting an outcome in the region if $i$ is true.

- We aim for low size and high power, but typically have to trade them off.

- Helpful pictures:

  - Developing the theory

    - Domination

      - $R$ dominates $R'$ just if $R$ has lower size and higher power, or lower size and the same power, or the same size and higher power.

      - Don’t choose a dominated rejection region.

      - What more can we say?

2. Developing the theory

Domination

- $R$ dominates $R'$ just if $R$ has lower size and higher power, or lower size and the same power, or the same size and higher power.

- Don’t choose a dominated rejection region.

- What more can we say?
A rule of thumb:

Choose a maximum size. Among undominated rejection regions with at most that size, choose one with the highest power.

Likelihood ratio tests

Working out the sizes and powers of all possible rejection regions is hard.

Can we save ourselves some work? Yes.

The likelihood ratio $L(x)$ of an outcome is its likelihood (or density) according to $i$ divided by its likelihood (or density) according to $h$.

What are the likelihood ratios in the discrete example?

Here’s a plot for the continuous example:

![Likelihood ratio for N(0, 1) against N(2, 1)](image)

Define a class of rejection regions: $R_\zeta := \{ x : L(x) > \zeta \}$.

The famous Neyman-Pearson Lemma:

Rejection regions of the form $R_\zeta$ aren’t dominated.

The lemma leads to a simpler rule of thumb:

Choose a maximum size. Among rejection regions of the form $R_\zeta$ with at most that size, choose one with the highest power.

Helpful pictures. (Orange dots are dominated. Big dots are of form $R_\zeta$ for some $\zeta$.)
Fancier stuff

▷ Mixed tests:
  - What if there’s no rejection region of the desired size?
  - Instead of a rejection region, which is a set of outcomes, use a rejection function, which is a function from outcomes to probabilities.
  - If we observe \( x \in X \), with probability \( f(x) \) we reject \( h \) and accept \( i \).
  - We can always find a rejection function of the desired size.

▷ Composite hypotheses:
  - What if we want to decide whether the coin is fair or heads-biased?
  - Think about the hypotheses pairwise: fair vs. .6-heads biased, fair vs. .7 heads-biased, ...
  - The concepts of size and power still apply pairwise, but there’s more to trade off.

3. Evaluating the methodology

How does it do in the long-run?

▷ Neyman and Pearson’s own suggestion:

Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong. [...] Such a rule tells us nothing as to whether in a particular case \( H \) is [true or false]. But it may often be proved that if we behave according to such a rule, then in the long run we shall reject \( H \) when it is true not more, say, than once in a hundred times, and in addition we may have evidence that we shall reject \( H \) sufficiently often when it is false.

1933, p. 291
Why aim for low size and high power?

▷ "A low probability of rejecting what is true; a high probability of rejecting what is false."

▷ But things may look different after the experiment than before:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>0.01</td>
<td>0.99</td>
</tr>
<tr>
<td>i</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>j</td>
<td>0.001</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Neyman-Pearson vs. Bayesians

▷ Bayesians use priors; N-P don’t.

▷ N-P reject and accept; Bayesians update credences and work out expected utilities.

▷ Is N-P easier to put into practice than Bayesianism?

▷ Is there enough common ground for one side to convince the other?

4. Resources

▷ For a more detailed version of the handout, and to play around with the code I used to make the pictures, see my github: cosmo-grant.


