small sample proportion

- what if S-F condition not met?
- inference via simulation
- setting up a simulation assuming H_0 true

Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel
Duke University

Image source: http://commons.wikimedia.org/wiki/File:Paul_the_Octopus_picks_Spain_over_Germany_in_2010_World_Cup_semi-final_3.jpg

practice

Paul the Octopus predicted 8 World Cup games, and predicted them all correctly. Does this provide convincing evidence that Paul actually has psychic powers, i.e. that he does better than just randomly guessing?

$H_0: p = 0.5$

$H_A: p > 0.5$

$n = 8$

$\hat{p} = 1$

1. independence:

we can assume that his guesses are independent

2. sample size / skew: $8 \times 0.5 = 4 \rightarrow$ not met

distribution of sample proportions cannot be assumed to be nearly normal

revisit: inference via simulation

- the ultimate goal of a hypothesis test is a p-value
  - p-value = P(observed or more extreme outcome | **H_0 true**)
- devise a simulation scheme that assumes the null hypothesis is true
- repeat the simulation many times and record relevant sample statistic
- calculate p-value as the proportion of simulations that yield a result favorable to the alternative hypothesis

Paul the Octopus predicted 8 World Cup games, and predicted them all correctly. Does this provide convincing evidence that Paul actually has psychic powers, i.e. that he does better than just randomly guessing?

$H_0 : p = 0.5$

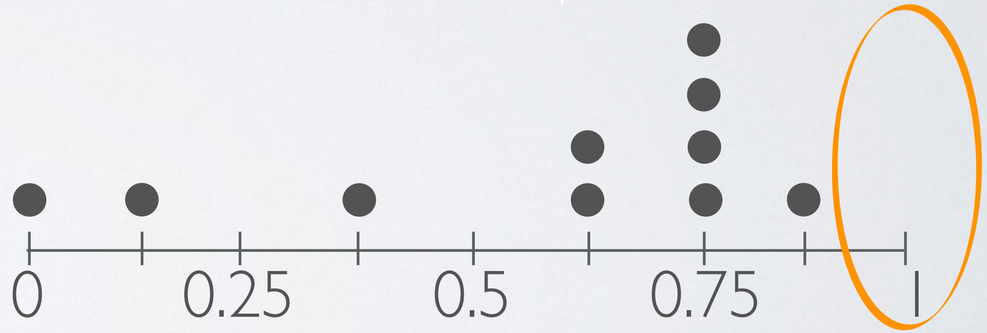
$H_A : p > 0.5$

- use a fair coin, and label head as success (correct guess)
- one simulation: flip the coin 8 times and record the proportion of heads (correct guesses) $\hat{p}_{sim}$
- repeat the simulation many times, recording the proportion of heads at each iteration $\hat{p}_{sim,1}, \hat{p}_{sim,2}, \cdots, \hat{p}_{sim,N}$
- calculate the percentage of simulations where the simulated proportion of heads is at least as extreme as the observed proportion

									$\hat{p}$
simulation 1:	H	H	H	T	H	H	H	H	7 / 8 = 0.875
simulation 2:	T	H	H	T	H	T	T	T	3 / 8 = 0.375
simulation 3:	T	T	H	H	H	H	T	H	5 / 8 = 0.625
…									…
simulation 10:	T	H	T	H	H	H	H	H	6 / 8 = 0.75

p-value $= P(\hat{p}_{sim} \geq 1 \mid p = 0.5)$

$= 0$



R

```
> source("http://bit.ly/dasi_inference")
> paul = factor(c(rep("yes", 8), rep("no", 0)), levels = c("yes","no"))
> inference(paul, est = "proportion", type = "ht", method = "simulation",
success = "yes", null = 0.5, alternative = "greater")
Single proportion -- success: yes
Summary statistics: p_hat = 1 ;  n = 8
H0: p = 0.5
HA: p > 0.5
p-value =  0.0037
```