Statistical Hypothesis Testing – recap –

Thursday 1st December, 2016

Experimenting

perform an experiment - to discover something

setting	experiment	discovery
alarm system	check sensors	intruder
new treatment for illness	test treatment on people	treatment works
hacker attack detection	test network traffic	there is an attack
software testing	run tests	bug

Alarm System

discovery = intruder

	test output = negative	test output = positive
no intruder	true negative	false positive
intruder	false negative	true positive

New Treatment for Illness

discovery = treatment works

	test output = negative	test output = positive
treatment doesn't work	true negative	false positive
treatment works	false negative	true positive

Hacker Attack Detection

discovery = hacker attack

	test output = negative	test output = positive
no attack	true negative	false positive
attack	false negative	true positive

Software Testing

discovery = bug

	test output = negative	test output = positive
no bug	true negative	false positive
bug	false negative	true positive

If we know what a discovery looks like

Alarm system

sensor readings X

no intruder	$0 \leq X \leq 4$
intruder	$5 \leq X \leq 8$

decision rule $X \overset{?}{>} 4$

no	no intruder
yes	intruder

If we know what a discovery looks like

New Treatment

number of tails T

sick	$T > 1$
not sick	$T = 0$

decision rule (treatment works if) $T \overset{?}{\leq} 1$

no	doesn't work
yes	work

If we know what a discovery looks like

Hacker Attack Detection

average upload, idle computer U

no attack	$U < 30\,[Kb/s]$
attack	$U > 50\,[Kb/s]$

decision rule (attack if) $U \overset{?}{\geq} 30$

no	no attack
yes	attack

If we know what a discovery looks like

Software Testing

number of failed tests N_f

no bug	$N_f = 0$
bug	$N_f > 0$

decision rule (there is a bug if) $N_f \overset{?}{>} 0$

no	no bug
yes	bug

If we don't know what a discovery looks like

Alarm system

sensor readings X

no intruder	$0 \leq X \leq 4$

decision rule $X \overset{?}{>} 4$

no	no intruder
yes	intruder

If we don't know what a discovery looks like

Israeltech challenge

New Treatment

number of tails T

sick	$T > 1$

decision rule (treatment works if) $T \overset{?}{\leq} 1$

no	doesn't work
yes	work

If we don't know what a discovery looks like

Israeltech challenge

Hacker Attack Detection

average upload, idle computer U

no attack	$U < 30\,[Kb/s]$

decision rule (attack if) $U \overset{?}{\geq} 30$

no	no attack
yes	attack

If we don't know what a discovery looks like

Software Testing

number of failed tests N_f

no bug	$N_f = 0$

decision rule (there is a bug if) $N_f \overset{?}{>} 0$

no	no bug
yes	bug

Experimenting with Distributions

- $N_1 = \mathcal{N}(\mu, \sigma^2)$, μ_0 constant
- Y fair coin

setting	experiment	discovery
is X not sampled fron N_1?	?	X not sampled from N_1
is Y unfair?	?	Y is unfair
is $\mu > \mu_0$?	?	$\mu > \mu_0$
X and Y are dependent?	?	X and Y are dependent

Is X not sampled from N_1

discovery $=$ X not sampled from N_1

	test output $=$ negative	test output $=$ positive
$X \sim N_1$	true negative	false positive
$X \nsim N_1$	false negative	true positive

Is Y unfair?

discovery $=$ Y is unfair

	test output $=$ negative	test output $=$ positive
Y is fair	true negative	false positive
Y is unfair	false negative	true positive

Is $\mu > \mu_0$?

discovery $= \mu > \mu_0$

	test output $=$ negative	test output $=$ positive
$\mu \leq \mu_0$	true negative	false positive
$\mu > \mu_0$	false negative	true positive

Are X and Y dependent?

discovery $=$ X and Y are dependent

	test output $=$ negative	test output $=$ positive
indpendent	true negative	false positive
dependent	false negative	true positive

If we know what a discovery looks like

Israeltech challenge

Is X not sampled from N_1?

X values

sampled from N_1	$X \sim N_1$
sampled from $N_2 = \mathcal{N}\left(\mu_2, \sigma^2\right)$	$X \sim N_2$

decision rule (MLE) $|X - \mu|^2 \overset{?}{>} |X - \mu_2|^2$

no	sampled from N_1
yes	not sampled from N_1

If we know what a discovery looks like

Is Y unfair?

coin tosses $Y_1, \ldots, Y_n$, $[\sum_{i=1}^{n} Y_i] = c$

fair	$P\left([\sum_{i=1}^{n} Y_i] = c\right) = \binom{n}{c} 0.5^n$
$P(Y = 1) = 0.8$	$P\left([\sum_{i=1}^{n} Y_i] = c\right) = \binom{n}{c} 0.8^c 0.2^{n-c}$

decision rule (MLE) $\binom{n}{c} 0.8^c 0.2^{n-c} > \binom{n}{c} 0.5^n$

no	fair
yes	unfair

If we know what a discovery looks like

Israeltech challenge

Is $\mu > \mu_0$?

sample mean $\overline{X} = \sum_{i=1}^{n} X_i/n$

$\mu = \mu_0$	$\overline{X} \sim \mathcal{N}\left(\mu_0, \sigma^2/n\right)$
$\mu = \mu_1 > \mu_0$	$\overline{X} \sim \mathcal{N}\left(\mu_1, \sigma^2/n\right)$

decision rule (MLE) $P\left(\overline{X} \mid \mu_1\right) > P\left(\overline{X} \mid \mu_0\right)$

no	$\mu = \mu_0$
yes	$\mu > \mu_0$

Logical Acrobatics

If we **know** what a discovery looks like

- we know the "normal" explanation — hypothesis
- we know the alternative explanation — hypothesis
- decision rule:

probability of results given the "normal" explanation $<$ probability of results given the alternative explanation

inaccurate

Logical Acrobatics

If we **know** what a discovery looks like

- we know the "normal" explanation — hypothesis
- we know the alternative explanation — hypothesis
- decision rule:

improbability of results given the "normal" explanation > improbability of results given the alternative explanation

inaccurate

Logical Acrobatics

Israeltech challenge

If we **don't know** what a discovery looks like

- we know the "normal" explanation — hypothesis
- we don't know the alternative explanation — hypothesis
- decision rule:

improbability of results given the "normal" explanation $>$ threshold th'

inaccurate

Logical Acrobatics

If we **don't know** what a discovery looks like

- we know the "normal" explanation hypothesis
- we don't know the alternative explanation hypothesis
- decision rule:

probability of results given the "normal" explanation $<$ threshold th

inaccurate

Probability of False Positive

Israeltech challenge

a false positive

- "normal" explanation - H_0

reject "normal" explanation <u>but</u> "normal" explanation is true

Probability of False Positive

a false positive

- "normal" explanation - H_0

reject "normal" explanation	<u>but</u>	"normal" explanation is true

- alarm system

reject - "no intruder"	<u>but</u>	there is no intruder

Probability of False Positive

Israeltech challenge

a false positive

- "normal" explanation - H_0

reject "normal" explanation	<u>but</u>	"normal" explanation is true

- testing new treatment

reject - "treatment doesn't work"	<u>but</u>	treatment doesn't work

Probability of False Positive

a false positive

- "normal" explanation - H_0

reject "normal" explanation	<u>but</u>	"normal" explanation is true

- hacker attack detection

reject - "no attack"	<u>but</u>	there is no attack

Probability of False Positive

a false positive

- "normal" explanation - H_0

reject "normal" explanation	<u>but</u>	"normal" explanation is true

- software testing

reject - "no bug"	<u>but</u>	there is no bug

Probability of False Positive

a false positive

- "normal" explanation - H_0

reject "normal" explanation	<u>but</u>	"normal" explanation is true

- Is X not sampled from N_1?

reject - X sampled from N_1	<u>but</u>	X is sampled from N_1

double negative - board

Probability of False Positive

a false positive

- "normal" explanation - H_0

reject "normal" explanation	<u>but</u>	"normal" explanation is true

- is Y unfair?

reject - Y is fair	<u>but</u>	Y is fair

double negative

Probability of False Positive

a false positive

- "normal" explanation - H_0

  reject "normal" explanation __but__ "normal" explanation is true

- Is $\mu > \mu_0$?

  reject - $\mu \leq \mu_0$ __but__ $\mu \leq \mu_0$

Probability of False Positive

Israeltech challenge

a false positive

- "normal" explanation - H_0

reject "normal" explanation	<u>but</u>	"normal" explanation is true

- X and Y are dependent?

reject - X and Y are independent	<u>but</u>	X and Y are independent

Rejected Results

- we reject "normal" explanation H_0 for some results
  - testing treatment - 95% of patients cured
  - software testing - 8/10 tests fail
  - testing a coin - 9/10 heads
- we do not reject for others
  - testing treatment - 0.5% of patients cured
  - software testing - no tests fail
  - testing a coin - 6/10 heads

Rejected Results

- we reject "normal" explanation H_0 for some results
- we do not reject for others
- results we reject
  - are $\boxed{\text{possible}}$ given "normal" explanation H_0
  - just $\boxed{\text{unlikely}}$

Rejected Results

Israeltech challenge

- we reject "normal" explanation H_0 for some results
- we do not reject for others
- results we reject
  - are $\boxed{\text{possible}}$ given "normal" explanation H_0
  - just $\boxed{\text{unlikely}}$

$$\begin{array}{c}\text{probability of results we reject}\\ \text{given the "normal" explanation } H_0\end{array} = \begin{array}{c}\text{probability of}\\ \text{false positives}\end{array}$$

Rejected Results

- if we reject for
  - testing treatment - 95% of patients cured
  - software testing - 8/10 tests fail
  - testing a coin - 9/10 heads
- we will also reject for
  - testing treatment - 99% of patients cured
  - software testing - 9/10 tests fail
  - testing a coin - 10/10 heads

Rejected Results

- if we consider
  - testing a coin - 9/10 heads
- unlikely enough to reject "normal" explanation H_0
- we would also consider
  - 10/10 heads
  - 1/10 heads
  - 0/10 heads
- unlikely enough to reject

P-Value

- this value for 9/10 heads

$$f\left(9/10\right) = P\left(9/10 \mid H_0\right) + P\left(10/10 \mid H_0\right) + P\left(1/10 \mid H_0\right) + P\left(0/10 \mid H_0\right)$$

- is the P-Value for 9/10 heads

P-Value

- result X
- all results at least as unlikely (improbable)

$$R\left(X\right) = \left\{Z : P\left(Z \mid H_0\right) \leq P\left(X \mid H_0\right)\right\}$$

P-Value

- result X
- all results at least as unlikely (improbable)

$$R(X) = \{Z : P(Z \mid H_0) \leq P(X \mid H_0)\}$$

- P-Value

$$Pval(X) = \sum_{Z \in R(X)} P(Z \mid H_0)$$
$$= P(R(X) \mid H_0)$$

How do we choose a threshold?

we are willing to allow 5% false positives

- if $Pval\,(X) < th = 0.05$
  - reject "normal" explanation H_0 for result X
- else
  - do no reject

Does this guarantee a false positive rate?

- X_{th} the most likely result we consider unlikely

  multiple possible

  - example: 9/10 heads

Does this guarantee a false positive rate?

- X_{th} the most likely result we consider unlikely

  multiple possible

  - example: 9/10 heads
- we reject anything at least as unlikely as X_{th}
  - 9/10 heads, 10/10 heads, 1/10 heads, 0/10 heads

Does this guarantee a false positive rate?

- X_{th} the most likely result we consider unlikely

  multiple possible

  - example: 9/10 heads
- we reject anything at least as unlikely as X_{th}
  - 9/10 heads, 10/10 heads, 1/10 heads, 0/10 heads
- probability of false positive
  - probability of anything at least as unlikely as X_{th}

Does this guarantee a false positive rate?

- X_{th} the most likely result we consider unlikely

  multiple possible
  - example: 9/10 heads
- we reject anything at least as unlikely as X_{th}
  - 9/10 heads, 10/10 heads, 1/10 heads, 0/10 heads
- probability of false positive
  - probability of anything at least as unlikely as X_{th}
  - $Pval\left(X_{th}\right)$ P-Value of X_{th}

Does this guarantee a false positive rate?

- X_{th} the most likely result we consider unlikely

  multiple possible

  - example: 9/10 heads
- we reject anything at least as unlikely as X_{th}
  - 9/10 heads, 10/10 heads, 1/10 heads, 0/10 heads
- probability of false positive
  - probability of anything at least as unlikely as X_{th}
  - $Pval\left(X_{th}\right)$ P-Value of X_{th}

$$\boxed{Pval\left(X_{th}\right) = P\left(\text{reject } H_0 \mid H_0\right) < th}$$

How do we choose the null hypothesis H_0

- usually stands for "normality"
  - no surprise
  - no change
  - treatment doesn't work
  - no intruder
  - ...

How do we choose the null hypothesis H_0

- usually stands for "normality"
  - no surprise
  - no change
  - treatment doesn't work
  - no intruder
  - ...
- $\boxed{\text{always}}$ has a computable distribution
  - we can compute $P\left(X \mid H_0\right)$

Null hypothesis examples

Israeltech challenge

setting	experiment	null hypothesis
alarm system	check sensors	
new treatment for illness	test treatment on people	
hacker attack detection	test network traffic	
software testing	run tests	

Null hypothesis examples

setting	experiment	null hypothesis
alarm system	check sensors	no intruder
new treatment for illness	test treatment on people	
hacker attack detection	test network traffic	
software testing	run tests	

Null hypothesis examples

setting	experiment	null hypothesis
alarm system	check sensors	no intruder
new treatment for illness	test treatment on people	treatment doesn't work
hacker attack detection	test network traffic	
software testing	run tests	

Null hypothesis examples

setting	experiment	null hypothesis
alarm system	check sensors	no intruder
new treatment for illness	test treatment on people	treatment doesn't work
hacker attack detection	test network traffic	no attack
software testing	run tests	

Null hypothesis examples

setting	experiment	null hypothesis
alarm system	check sensors	no intruder
new treatment for illness	test treatment on people	treatment doesn't work
hacker attack detection	test network traffic	no attack
software testing	run tests	no bug

Null hypothesis examples

setting	experiment	null hypothesis
is X not sampled fron N_1?	sample X	
is Y unfair?	toss coin	
is $\mu > \mu_0$?	sample X, calculate sample mean	
X and Y are dependent?	sample (X, Y)	

Null hypothesis examples

setting	experiment	null hypothesis
is X not sampled fron N_1?	sample X	$X \sim N_1$
is Y unfair?	toss coin	
is $\mu > \mu_0$?	sample X, calculate sample mean	
X and Y are dependent?	sample (X, Y)	

Null hypothesis examples

setting	experiment	null hypothesis
is X not sampled fron N_1?	sample X	$X \sim N_1$
is Y unfair?	toss coin	Y is fair $p = 0.5$
is $\mu > \mu_0$?	sample X, calculate sample mean	
X and Y are dependent?	sample (X, Y)	

Null hypothesis examples

setting	experiment	null hypothesis
is X not sampled fron N_1?	sample X	$X \sim N_1$
is Y unfair?	toss coin	Y is fair $p = 0.5$
is $\mu > \mu_0$?	sample X, calculate sample mean	$\mu \leq \mu_0$ $\overline{X} \sim N\left(\mu_0, \sigma^2/n\right)$
X and Y are dependent?	sample (X, Y)	

Null hypothesis examples

setting	experiment	null hypothesis
is X not sampled fron N_1?	sample X	$X \sim N_1$
is Y unfair?	toss coin	Y is fair $p = 0.5$
is $\mu > \mu_0$?	sample X, calculate sample mean	$\mu \leq \mu_0$ $\overline{X} \sim N\left(\mu_0, \sigma^2/n\right)$
X and Y are dependent?	sample (X, Y)	X and Y are independent

Test statistic

test statistic T

what we compute P-Value for

$$Pval\left(X\right) = P\left(T\left(X\right) \mid H_0\right)$$

Test statistic

Israeltech challenge

test statistic T

what we compute P-Value for

example - Z-score

setting	experiment	test statistic*
is X not sampled fron N_1?	sample X	
is Y unfair?	toss coin	
is $\mu > \mu_0$?	sample X, calculate sample mean	
X and Y are dependent?	sample (X,Y)	

Test statistic

Israeltech challenge

test statistic T

what we compute P-Value for

example - Z-score

setting	experiment	test statistic*
is X not sampled fron N_1?	sample X	X
is Y unfair?	toss coin	
is $\mu > \mu_0$?	sample X, calculate sample mean	
X and Y are dependent?	sample (X, Y)	

Test statistic

test statistic T

what we compute P-Value for

example - Z-score

setting	experiment	test statistic*
is X not sampled fron N_1?	sample X	X
is Y unfair?	toss coin	$\sum_{i=1}^{n} Y_i$
is $\mu > \mu_0$?	sample X, calculate sample mean	
X and Y are dependent?	sample (X, Y)	

Test statistic

IsraelTech challenge

test statistic T

what we compute P-Value for

example - Z-score

setting	experiment	test statistic*
is X not sampled fron N_1?	sample X	X
is Y unfair?	toss coin	$\sum_{i=1}^{n} Y_i$
is $\mu > \mu_0$?	sample X, calculate sample mean	$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
X and Y are dependent?	sample (X, Y)	

Test statistic

Israeltech challenge

test statistic T

what we compute P-Value for

example - Z-score

setting	experiment	test statistic*
is X not sampled fron N_1?	sample X	X
is Y unfair?	toss coin	$\sum_{i=1}^{n} Y_i$
is $\mu > \mu_0$?	sample X, calculate sample mean	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
X and Y are dependent?	sample (X, Y)	$2\log\frac{P(X,Y)}{P(X)P(Y)}$ *

Questions?