Hypothesis Testing Introduction

Hypotheses

Assumptions about parameters

Simple & composite

Null & Alternative

Test

Accept or reject null

Preview

- ← Parameter estimation
  - Point estimates
  - Confidence intervals

- → Become the…

Grade credit

“Hypothesis Testing”

Hypotheses

Assumptions (statements) about parameters

Distribution

Population

A coin is biased

Average student GPA is < 3.0

Amazon average delivery time > 2 days

People tweet more on weekend

Men play more video games than women on average

Hypotheses Types

Simple — Parameter takes a single specific value

$\mu = 4.5$ $\sigma = 2.4$

Composite — Parameter takes one of several values

$\mu \in \{4.5, 6.3\}$ $\mu > \sigma$ $\sigma \in [4.5, 6.3)$

One-sided

$\mu \leq 2.3$ $\mu > 4.5$

Two-sided

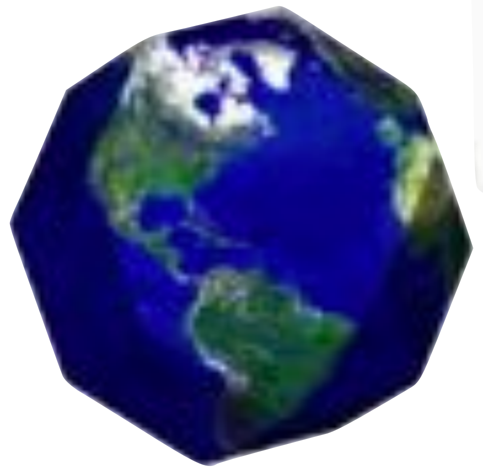
$\mu \leq 2.3$ or > 4.5 $\mu < 2.3$ or > 2.3 $\mu \neq 2.3$

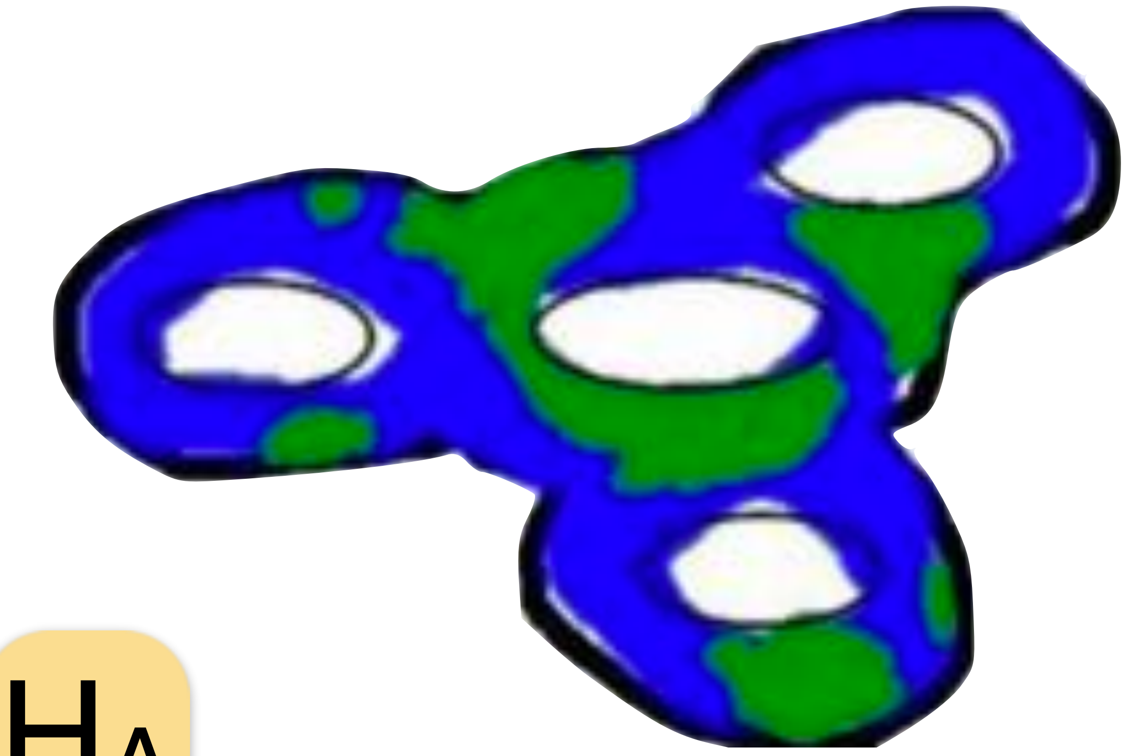
Null and Alternative Hypotheses

Often Assumption believed to be true

Null Status quo hypothesis H_0

Complementary view





Alternative Research hypothesis H_1 H_A

H_A often Complement or "one-side complement" of H_0

Simple H_0

Null

Alternative

Coin

Unbiased

$p_h=0.5$

Biased

$p_h \neq 0.5$

2-sided

Heads more likely

$p_h > 0.5$

1-sided

Gender equality average GPA

Same average GPA

Not exactly simple: $\{(x,x)\}$

Different average GPA

2-sided

Men’s average GPA is higher

1-sided

One-Sided H_0

Null | Alternative

Smartphones
iOS x Android

$\geq$ 60% use iOS

< 60% of phones use iOS

Checkout
Self x Cashier

Not exactly one sided: $\{(x,y): x<y\}$

Self checkout faster

Self checkout slower

How to Test

Design experiment

Gather data

Data consistent with null hypothesis?

Reject null in favor of alternative

Do not reject null

Equivalently

Strong evidence for alternative hypothesis?

Conservative

Reject null (status quo)

Only if strong evidence against it

Two analogies

Test v. Trial

Hypothesis test

Strong evidence for alternative hypothesis?

✅ → Reject null in favor of alternative

❌ → Do not reject null

Court trial

Strong incriminating evidence?

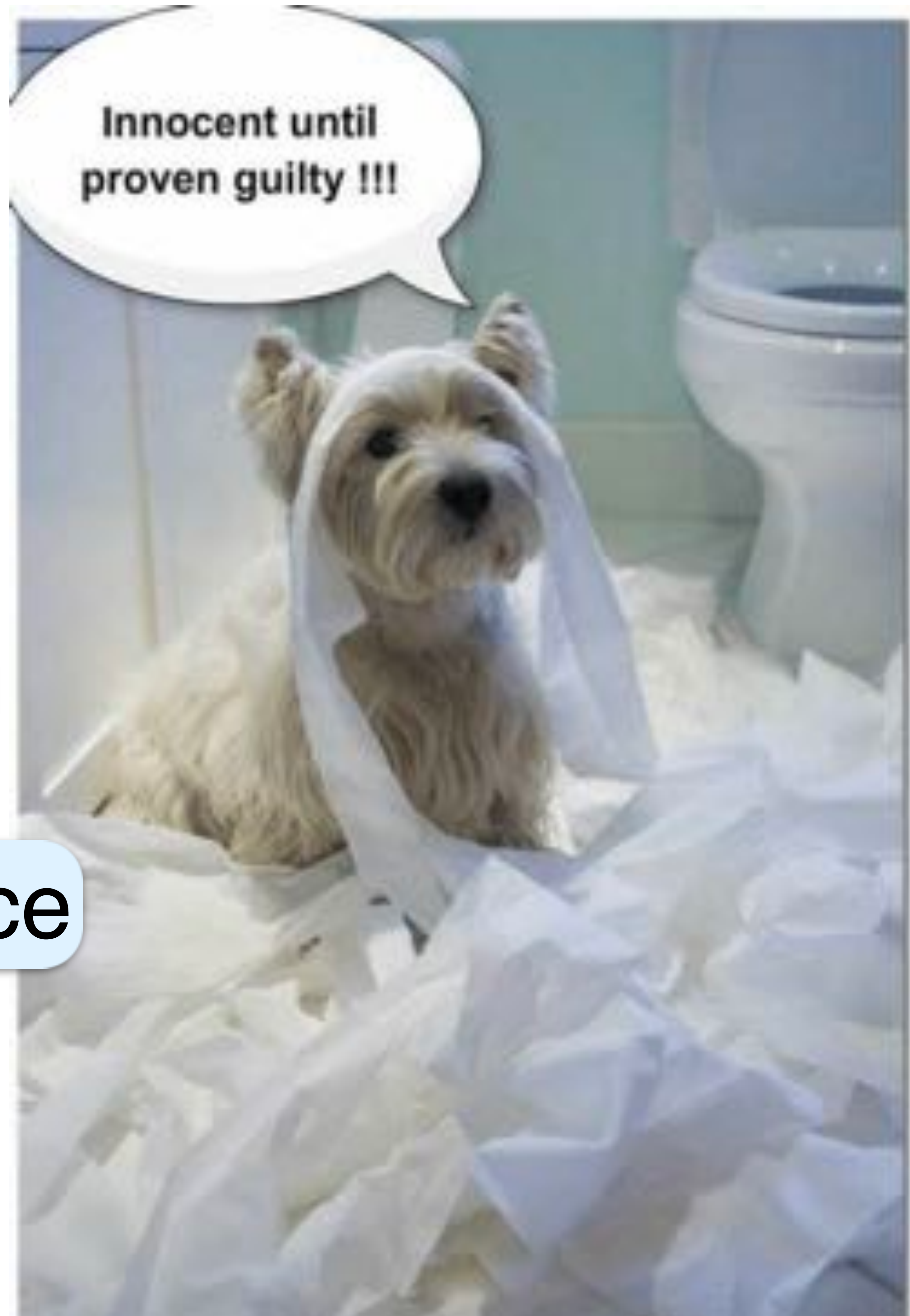
✅ → Guilty

❌ → Not guilty

Innocence = Null

Presumed innocence

Rejected only by strong evidence



Test vs. Myth

Hypothesis test

Strong evidence for alternative hypothesis?

✅ → Reject null in favor of alternative

❌ → Do not reject null

Strong evidence for myth?

✅ → Accept

❌ → Keep default belief

Design a test

Testing Hypotheses

Design Experiment   Test

Define numerical outcome   T   Test statistic   Related to hypothesis

Determine distribution of T under H_0 $P_{H_0}(T=t)$

Observe data   Calculate value t of the test statistic T

Large

$P_{H_0}(t)$

Small
t towards H_A

H_0 consistent with data   Do not reject H_0 Accept H_0

H_0 inconsistent with data   Reject H_0 in favor of H_A

Intuitive → Formal

Coin Bias, 1-Sided H_A

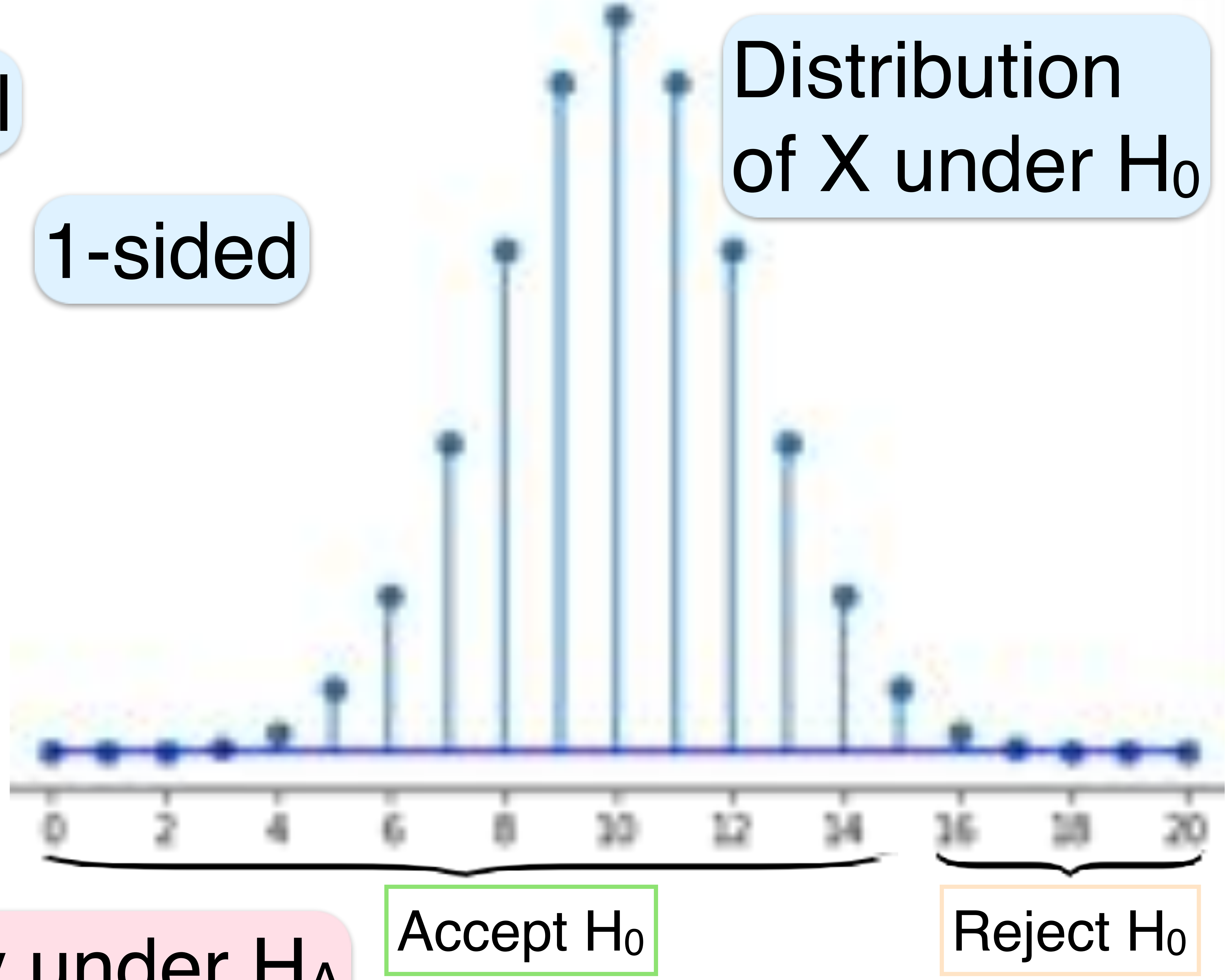
H_0 | Unbiased | $p_h = 0.5$ | Simple null

H_A | Biased towards heads | $p_h > 0.5$ | 1-sided

Test

Test statistic | Number of heads | X

Distribution of X under H_0



Intuitive

$X \geq 16$ | Unlikely under H_0 | More likely under H_A

Reject null in favor of H_A

$X < 15$ | Do not reject null

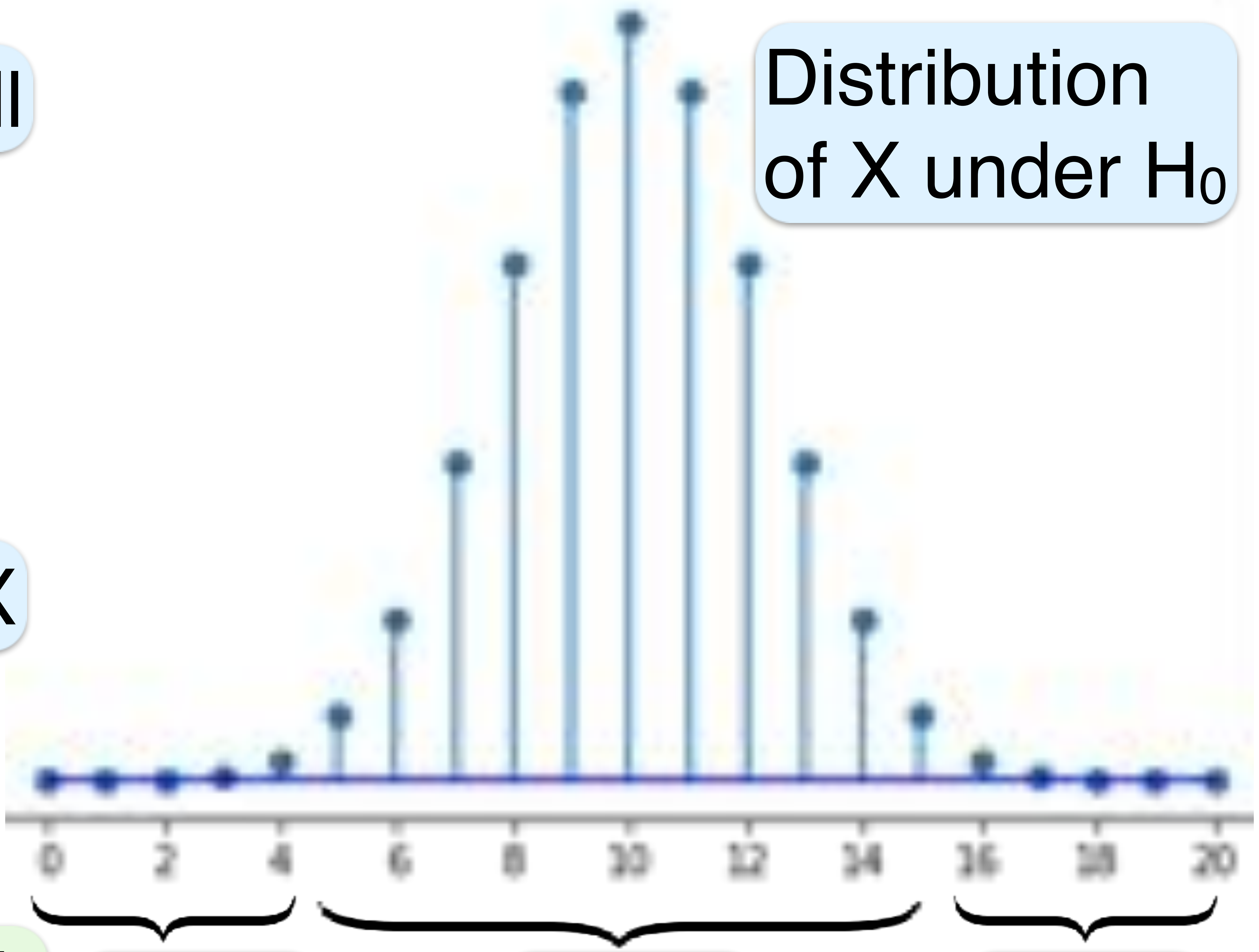
Coin Bias, 2-Sided H_A

H_0 Unbiased $p_h = 0.5$ Simple null

H_A Biased $p_h \neq 0.5$ 2-sided

Test 20X 

Test statistic Number of heads X



Distribution of X under H_0

Intuitively

$5 \leq X \leq 15$ Do not reject H_0 Accept H_0

Reject H_0 Accept H_0 Reject H_0

Otherwise Reject H_0 in favor of H_A

Hypothesis Testing Introduction

Assumptions about parameters

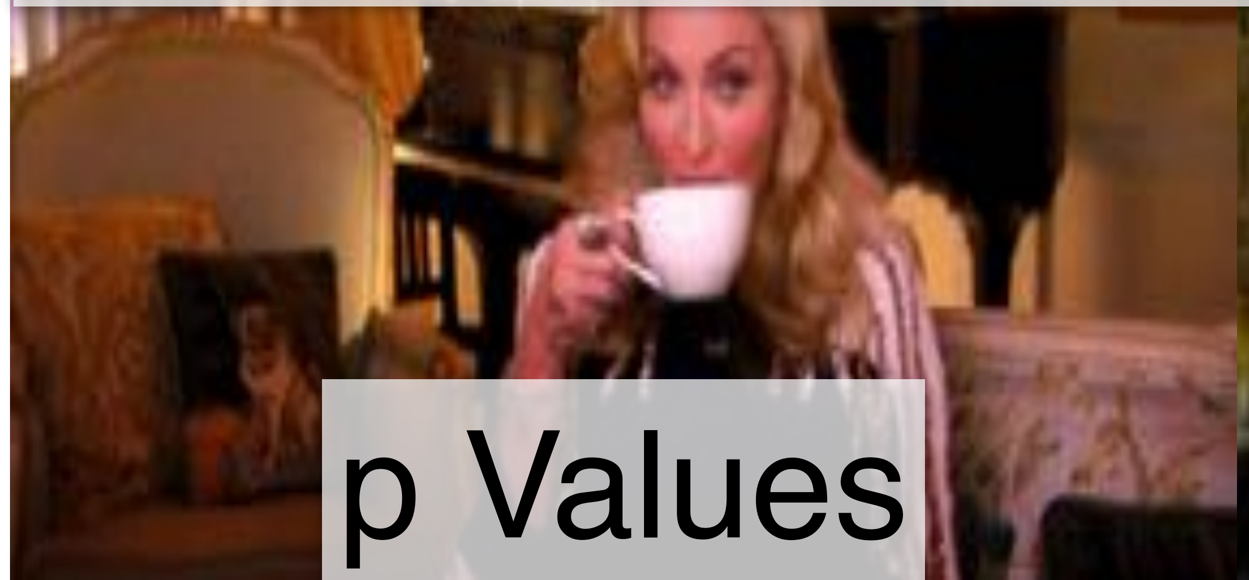
Simple & composite

Null & Alternative

Test

Accept or reject null

Hypothesis Testing



p Values

Hypothesis Testing - p Values

More systematic & precise

Significance level

Critical values

p values

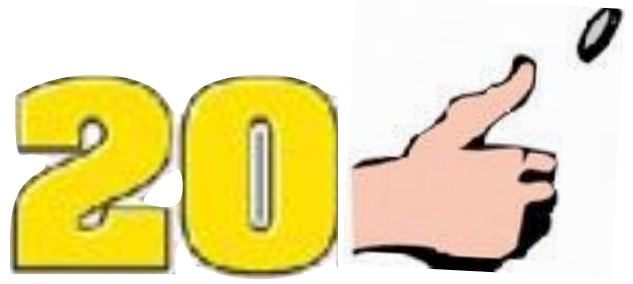
One- two- sided alternatives


Found something we like
Put a number on it
5%
Just need

Coin Bias

$H_0 : p_h = 0.5$ — Unbiased

$H_A : p_h > 0.5$ — Heads more likely

Data 

X : #  — Test Statistic

Type-I Error

H_0 : Coin unbiased

We declare H_A : heads more likely

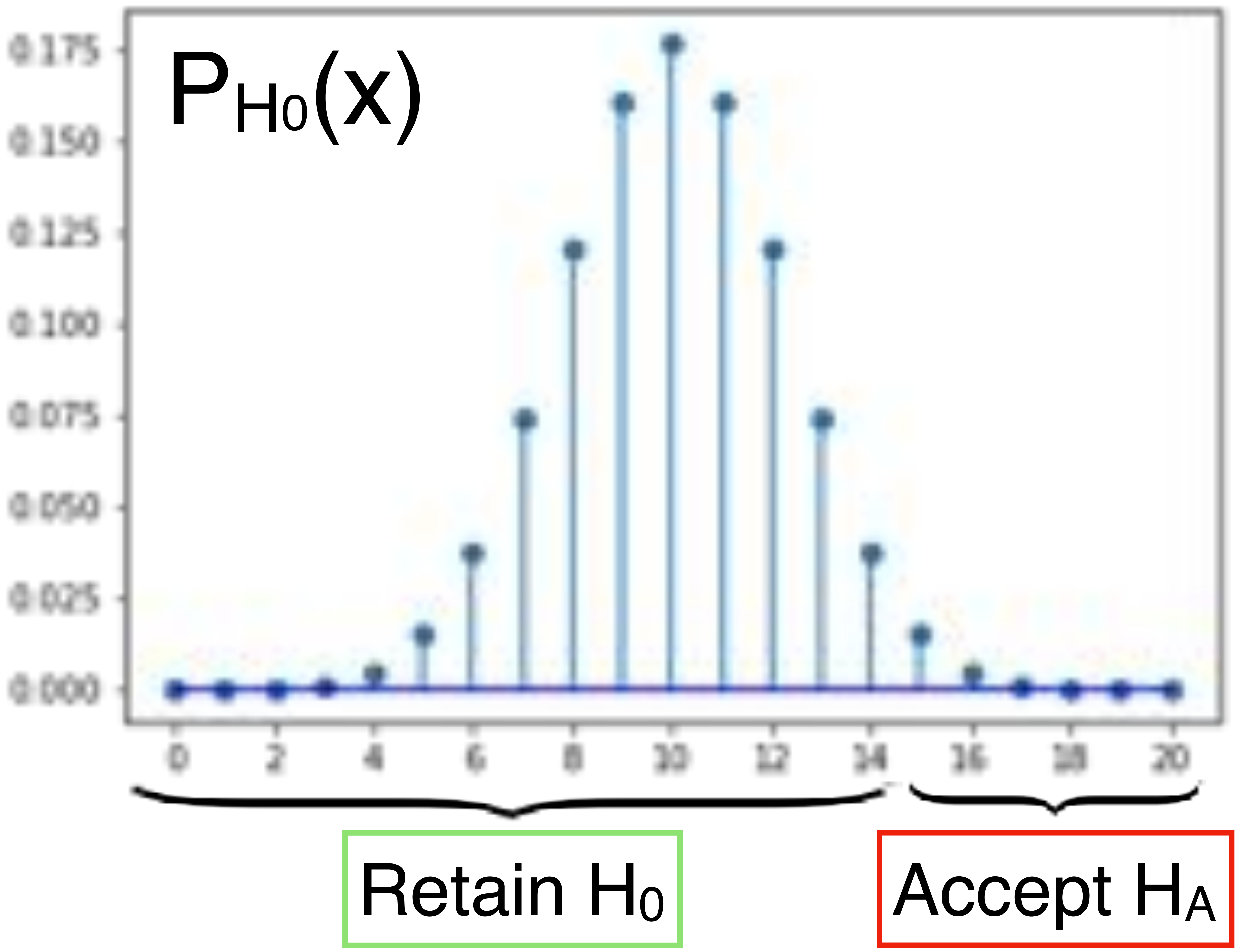
Under H_0

$X \sim B_{0.5,20}$

$P_{H_0}(x)$

← Eyeballed

→ More systematic



Nomenclature

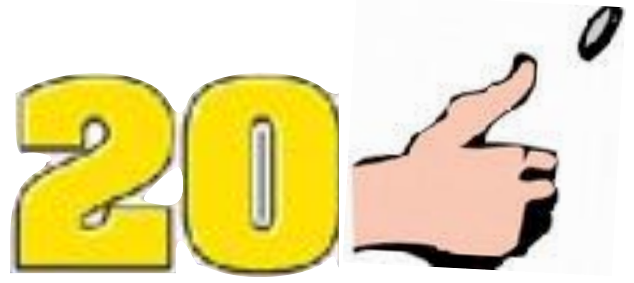
Complicated word for terminology | Apropos | Appropriate

H_A | Reject H_0 in favor of H_A | Reject H_0 | Accept H_A

H_0 | Do not reject H_0 | Data not significant | H_0 plausible | ~~Accept H_0~~ | Retain H_0

"Reluctantly retain"

Why verbal gymnastics? | Test asymmetric



X=12 | Accept H_0?

What if $p_h = 0.6$? | 0.55? | Better explain 12 than $p_H = 0.5$

Don't know that H_0 true | Just not enough data to reject it

Significance Level

Reject null (status quo) hypothesis H_0 only if strong evidence for alternative H_A

Precise probabilistic formulation

Significance level α Typically 5% 1%

Tiny

If H_0 is true, accept H_A with probability $\le \alpha$

$P_{H_0}(\text{ accept } H_A) \le \alpha$

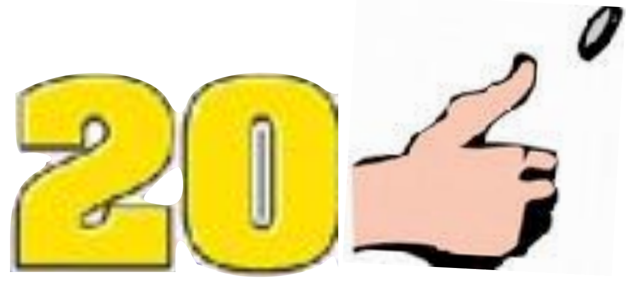
Type-I error

Two methods

Critical Values

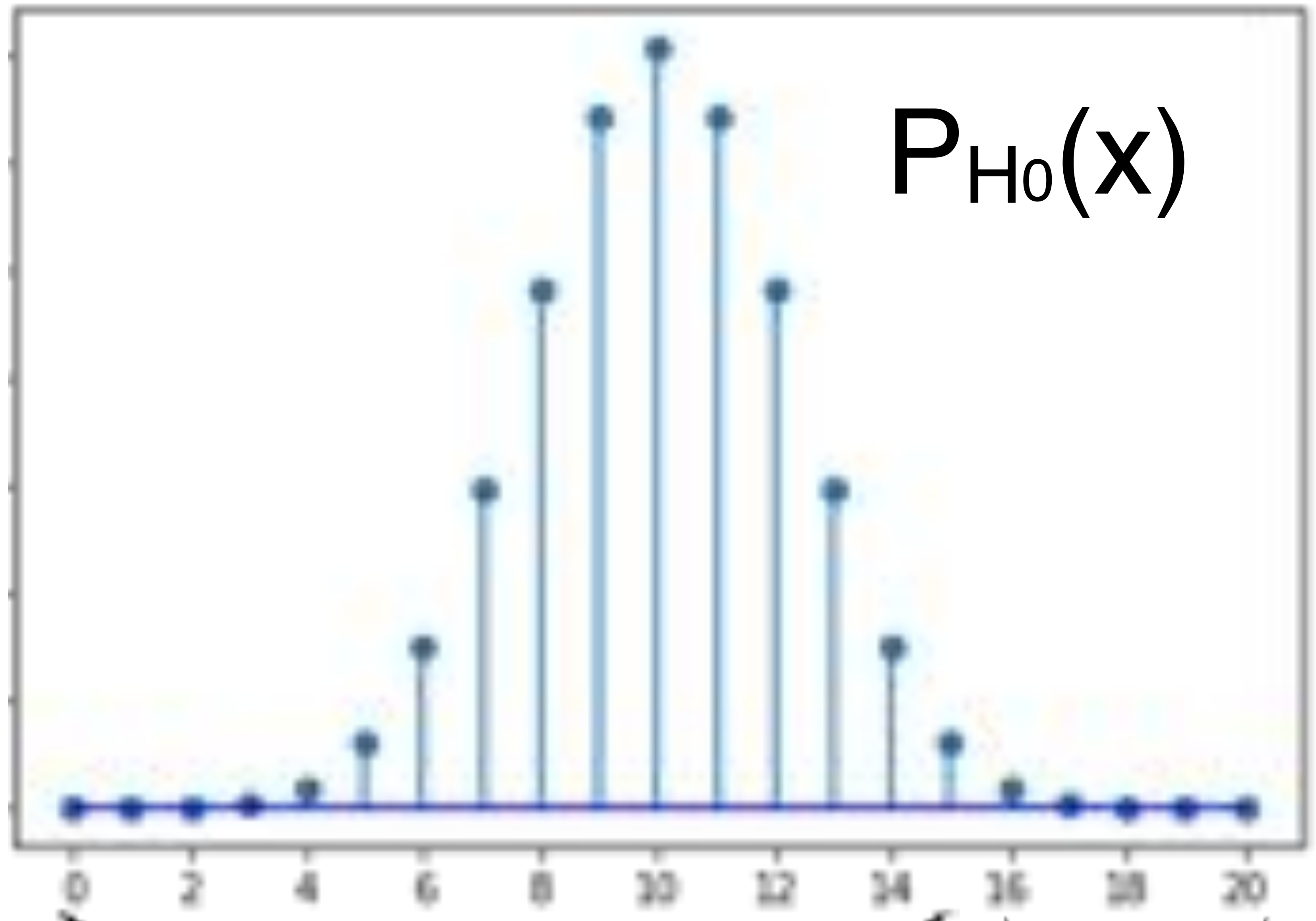
p Values

Critical-Value

$H_0 : p_h = 0.5$ $H_A : p_h > 0.5$ Data  X : # 

H_0 $X \approx 10$ If accept H_A when $X =$ 16 → 17 → 18 → 19 → 20

Threshold   Critical value x_α



Retain H_0 Accept H_A

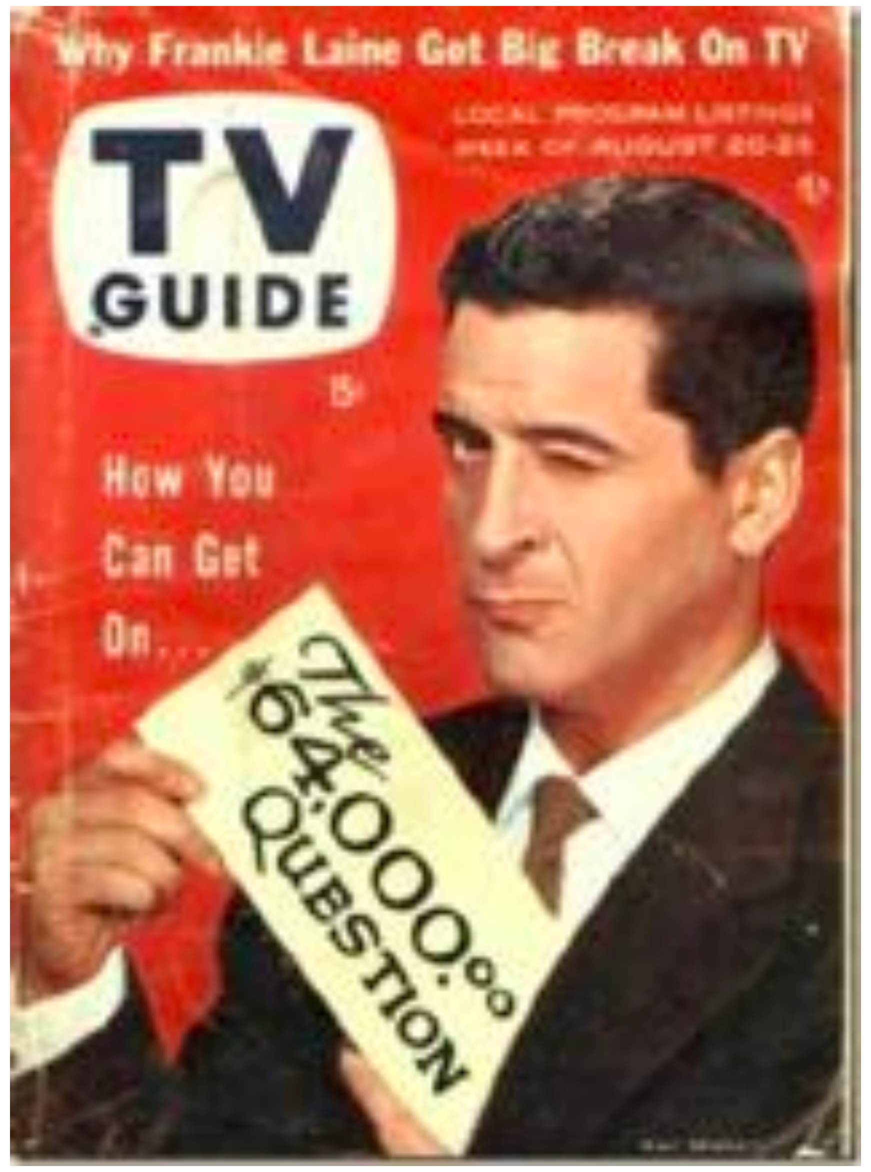
critical value x_α

X: $\geq x_\alpha$ → Accept H_A

X: $< x_\alpha$ → Retain H_0

x_α ← Significance level α

Upper bound on $P_{H_0}(\text{accept } H_A)$



What is x_α ?

Finding x_α

$H_0 : p_h = 0.5$ $H_A : p_h > 0.5$ Data

X : #

Critical value x_α $X < x_\alpha \rightarrow$ Retain H_0 $X \geq x_\alpha \rightarrow$ accept H_A

α Significance level 5% 1%

Type-I error

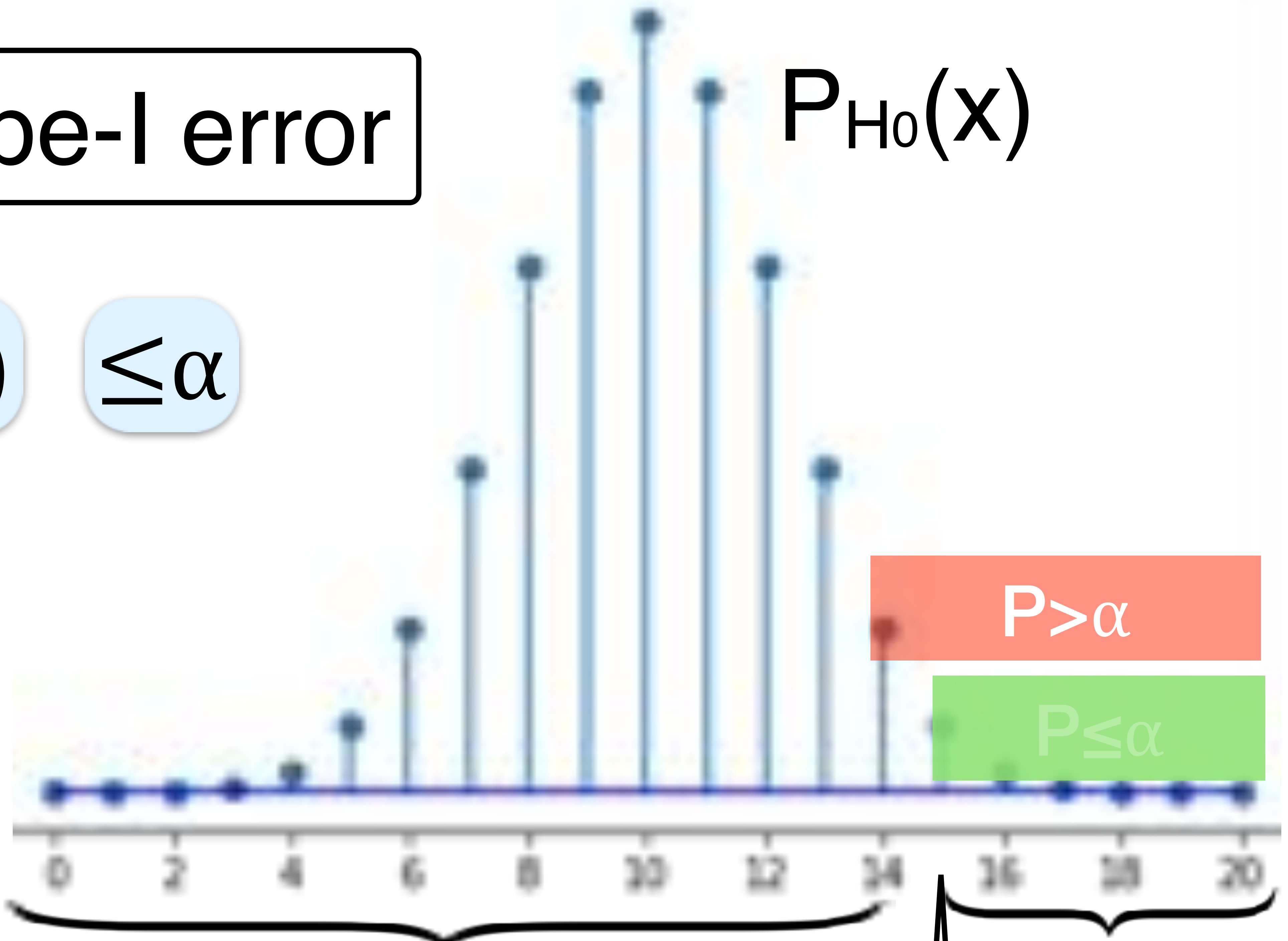
$P_{H_0}(X \geq x_\alpha) =$ P_{H_0} (falsely accept H_A) $\leq \alpha$

Need $P_{H_0}(X \geq x_\alpha) \leq \alpha$

x_α large Almost never declare H_A

Smallest x such that $P_{H_0}(X \geq x) \leq \alpha$

$P_{H_0}(x)$



Retain H_0 x_α Accept H_A

Example: $x_{5\%}$ and $x_{1\%}$

$H_0 : p_h = 0.5$

$H_A : p_h > 0.5$

Data 

X : # 

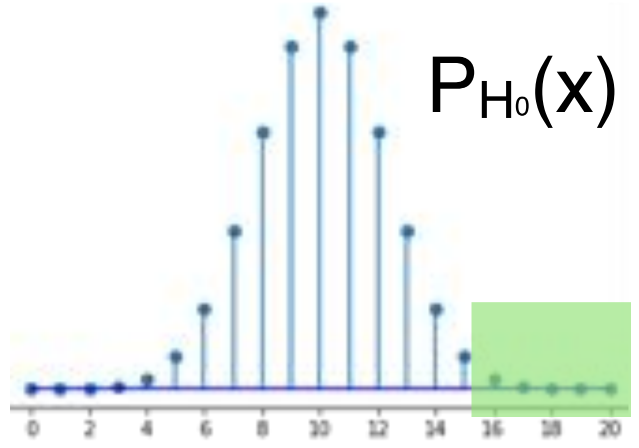
Significance $\alpha = 5\%$

Critical value $x_{5\%}$ Smallest x $P_{H_0}(X \geq x) \leq 5\%$

#heads x	$P_{H_0}(x)$	%	$P_{H_0}(X \geq x)$
20	$\binom{20}{20}\frac{1}{2^{20}}$	0.0001%	0.0001%
19	$\binom{20}{19}\frac{1}{2^{20}}$	0.0019%	0.002%
18	$\binom{20}{18}\frac{1}{2^{20}}$	0.0181%	0.0201%
17	$\binom{20}{17}\frac{1}{2^{20}}$	0.1087%	0.1288%
16	$\binom{20}{16}\frac{1}{2^{20}}$	0.4621%	0.5909%
15	$\binom{20}{15}\frac{1}{2^{20}}$	1.4786%	2.0695%
14	$\binom{20}{14}\frac{1}{2^{20}}$	3.6964%	5.7659%

P ≤ 5 %

P > 5 %



$P_{H_0}(X \geq 15) \approx 2.07\% \leq 5\%$

$P_{H_0}(X \geq 14) \approx 5.77\% > 5\%$

Accept H_A

$\boxed{x_{1\%} = 16}$

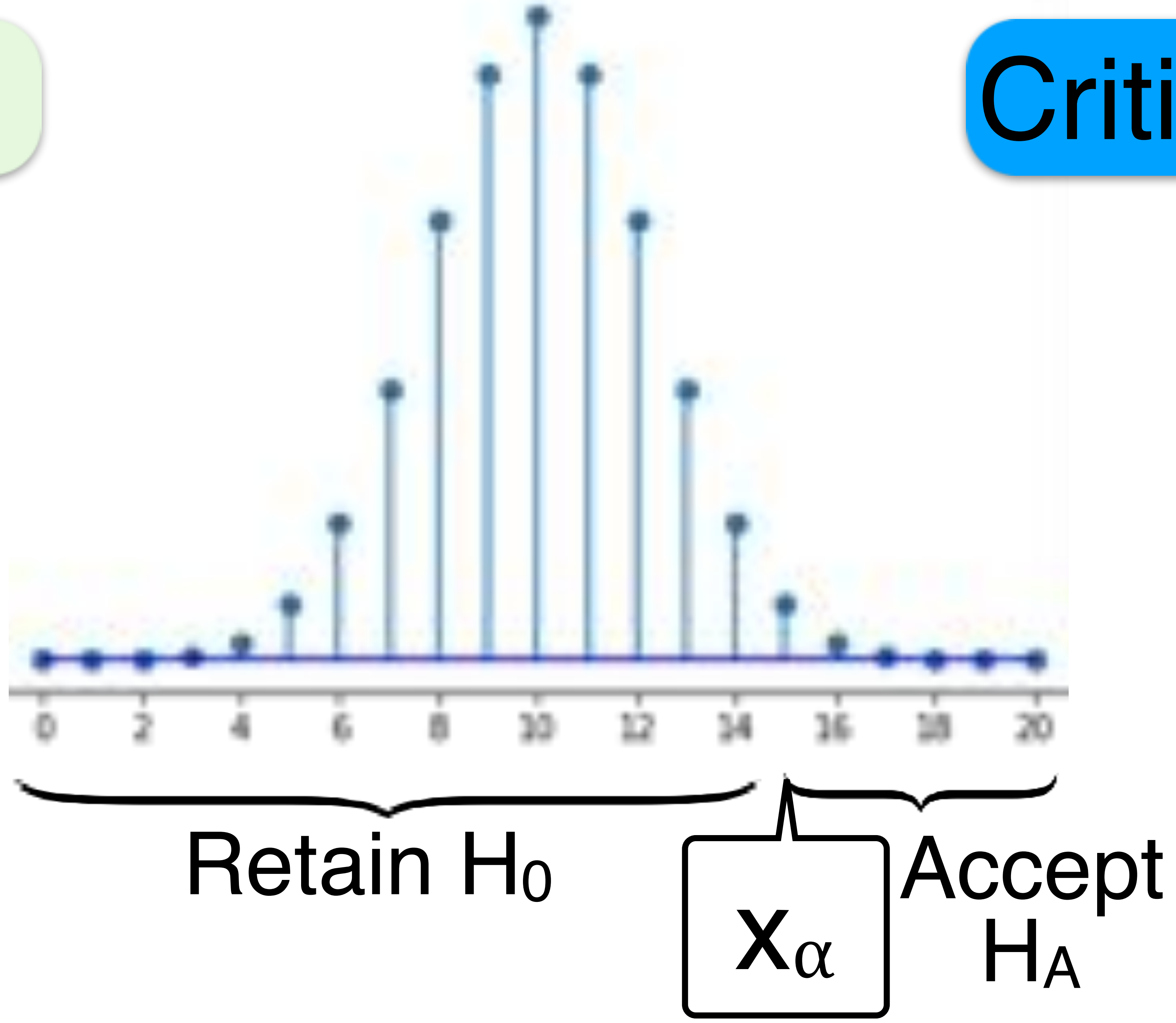
$\boxed{x_{5\%} = 15}$

Retain H_0

Room for Improvement

Significance level | α | 5%

Critical value | x_α | 15



X $< x_\alpha$ Retain H_0

X $\geq x_\alpha$ Accept H_A

Practically | Observe one value x of X | $X = 13$ | Retain H_0 or accept H_A

Can we..

Get around finding smallest x for accepting H_A?

Find a rule just for X itself?

p (probability) Values

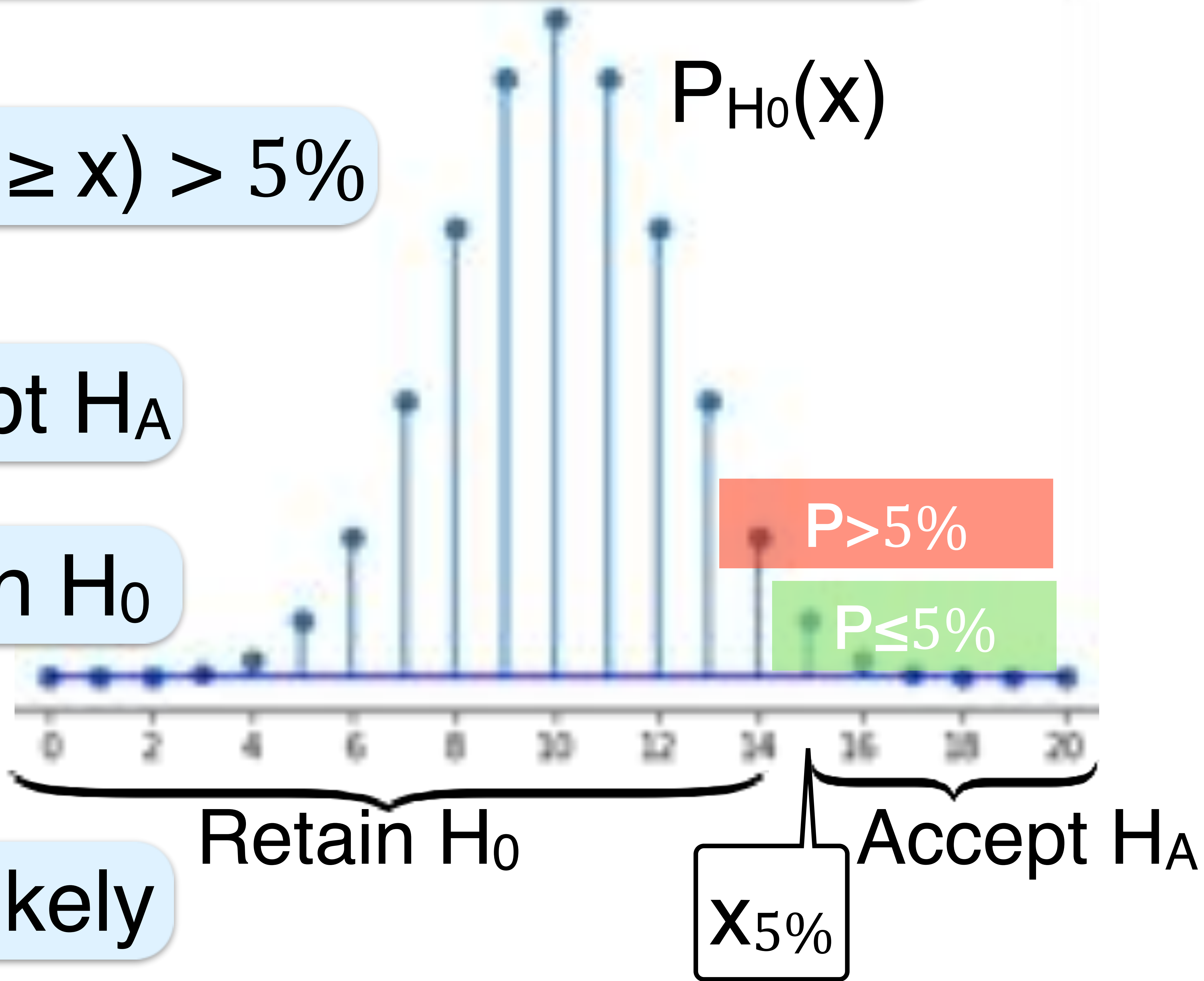
Critical value | $x_{5\%}$ | 15 | Smallest x | $P_{H_0}(X \geq x) \leq 5\%$

x

- $\geq x_{5\%}$ Accept H_A $\quad x \geq x_{5\%} \leftrightarrow P_{H_0}(X \geq x) \leq P_{H_0}(X \geq x_{5\%}) \leq 5\%$
- $< x_{5\%}$ Retain H_0 $\quad x < x_{5\%} \leftrightarrow P_{H_0}(X \geq x) > 5\%$

p value of x | $P_{H_0}(X \geq x)$

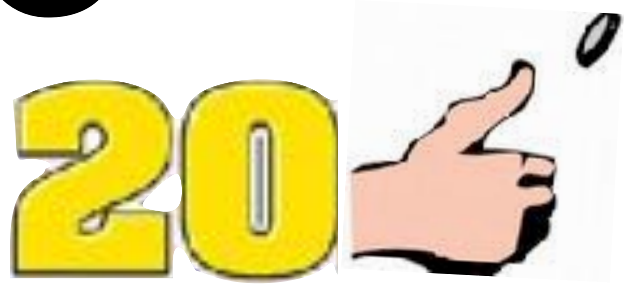
- $\leq 5\%$ | $x \geq x_{5\%}$ | Accept H_A
- $> 5\%$ | $x < x_{5\%}$ | Retain H_0



Same H_0 and H_A regions as before!

Intuitively | P under H_0 small $\rightarrow$ H_A more likely

p Values Example

$H_0 : p_h = 0.5$ $H_A : p_h > 0.5$ Data  X : # 

Significance $\alpha = 5\%$ **Critical value** $x_{5\%} = 15$ $X \geq 15 \rightarrow H_A$ $X < 15 \rightarrow H_0$

# heads x	$P_{H_0}(x)$	%	$P_{H_0}(X \geq x)$	
20	$\binom{20}{20}\frac{1}{2^{20}}$	0.0001%	0.0001%	H_A
19	$\binom{20}{19}\frac{1}{2^{20}}$	0.0019%	0.002%	•
18	$\binom{20}{18}\frac{1}{2^{20}}$	0.0181%	0.0201%	•
17	$\binom{20}{17}\frac{1}{2^{20}}$	0.1087%	0.1288%	•
16	$\binom{20}{16}\frac{1}{2^{20}}$	0.4621%	0.5909%	•
15	$\binom{20}{15}\frac{1}{2^{20}}$	1.4786%	2.0695%	H_A
14	$\binom{20}{14}\frac{1}{2^{20}}$	3.6964%	5.7659%	H_0

$\alpha = 1\%$: 15 → H_0

p value | $\leq 5\%$ | $x \geq x_{5\%}$ | Accept H_A

$> 5\%$ | $x < x_{5\%}$ | Retain H_0

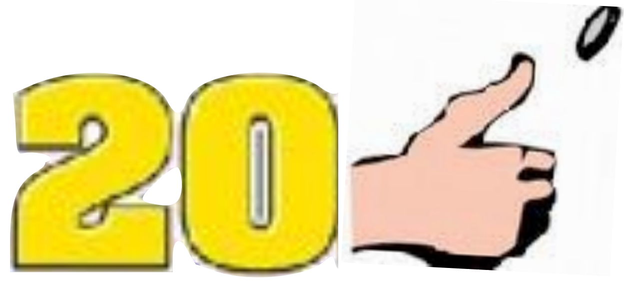
p value of x

Probability $X \geq x$

$\leq \alpha$ accept H_A

$> \alpha$ retain H_0

Opposite Alternative

$H_0 : p_h = 0.5$ $H_A : p_h < 0.5$ Data  X : # 

Significance level α 5%   1% P_{H_0} (falsely accept H_A) $\leq \alpha$

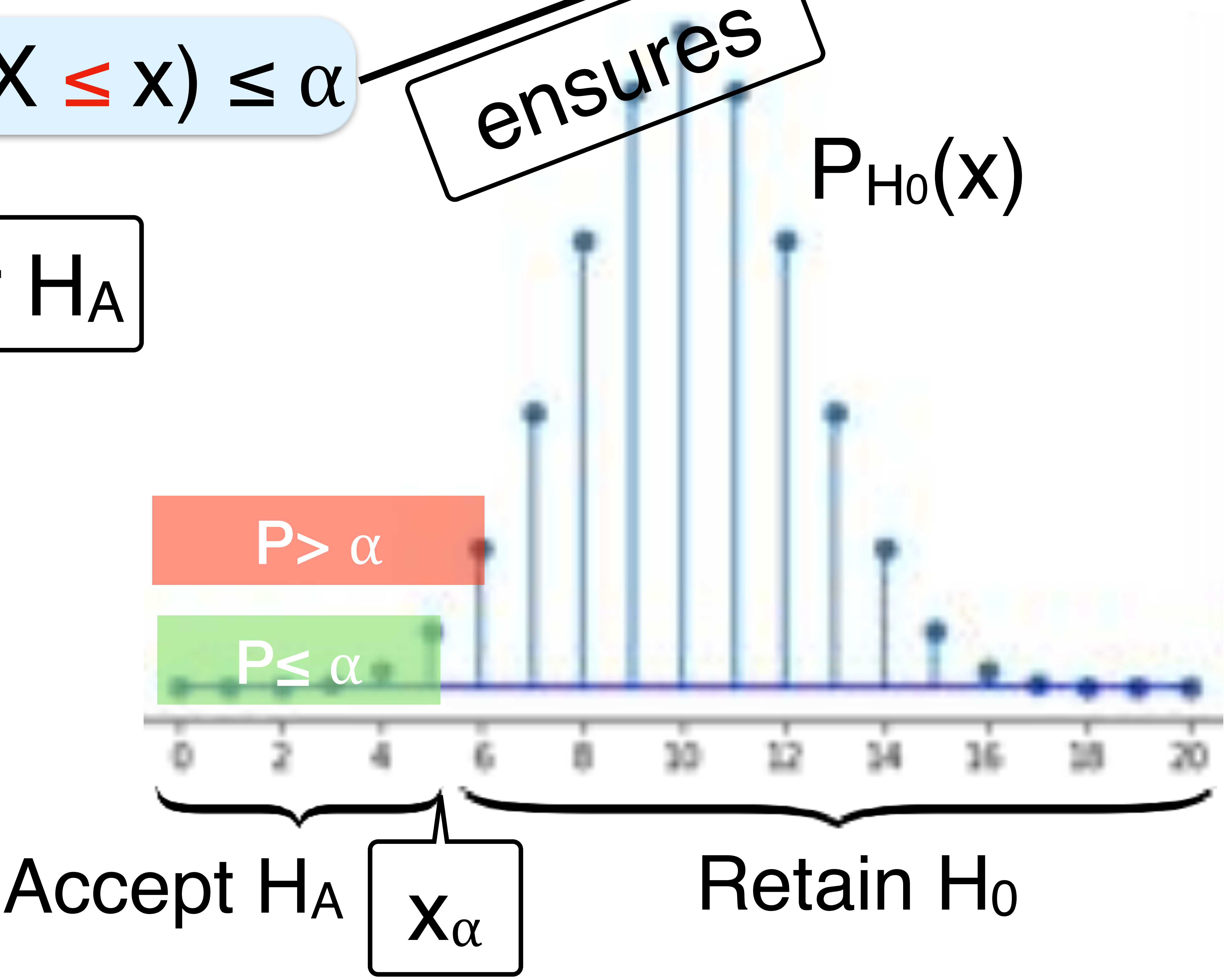
Critical value x_α Largest x s.t. $P_{H_0}(X \leq x) \leq \alpha$

ensures

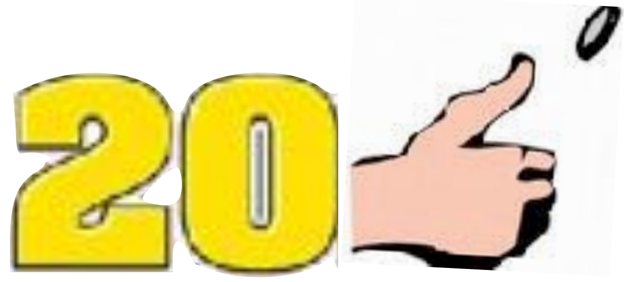
Small error under H_A

X
- $\leq x_\alpha$ Accept H_A
- $> x_\alpha$ Retain H_0

$\alpha = 5\%$ By symmetry $x_{5\%} = 5$



Using p Values

$H_0 : p_h = 0.5$ $H_A : p_h < 0.5$ Data  X : # 

Significance level $\alpha = 5\%$

# heads x	$P_{H_0}(x)$	%	$P_{H_0}(X \leq x)$
0	$\binom{20}{20}\frac{1}{2^{20}}$	0.0001%	0.0001%
1	$\binom{20}{19}\frac{1}{2^{20}}$	0.0019%	0.002%
2	$\binom{20}{18}\frac{1}{2^{20}}$	0.0181%	0.0201%
3	$\binom{20}{17}\frac{1}{2^{20}}$	0.1087%	0.1288%
4	$\binom{20}{16}\frac{1}{2^{20}}$	0.4621%	0.5909%
5	$\binom{20}{15}\frac{1}{2^{20}}$	1.4786%	2.0695%
6	$\binom{20}{14}\frac{1}{2^{20}}$	3.6964%	5.7659%

p value

$\leq 5\%$	$x \leq x_{5\%}$	Accept H_A
$> 5\%$	$x > x_{5\%}$	Retain H_0

H_A (x = 0)

•

•

•

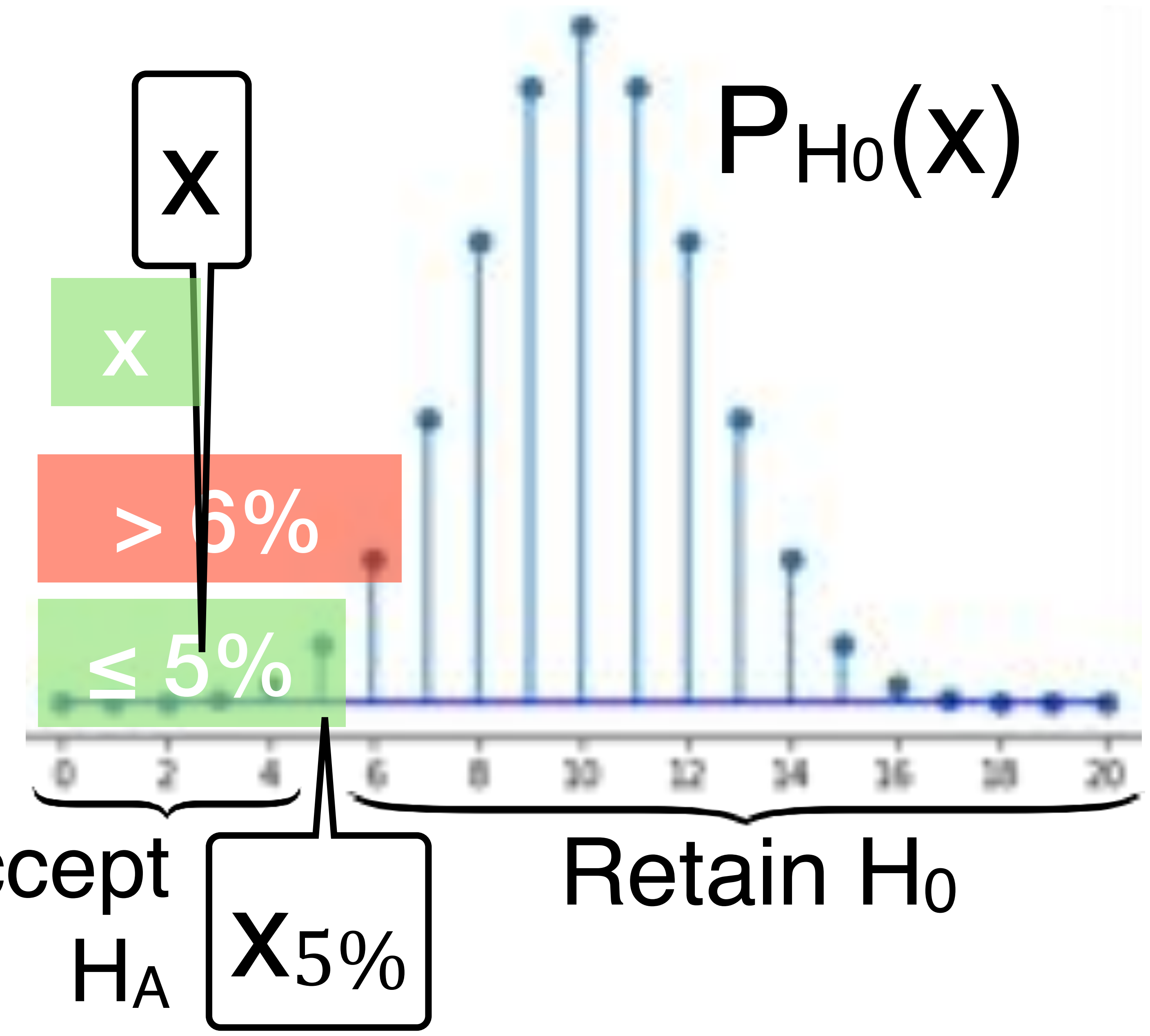
•

H_A (x = 5)

H_0 (x = 6)

$\alpha = 1\%$

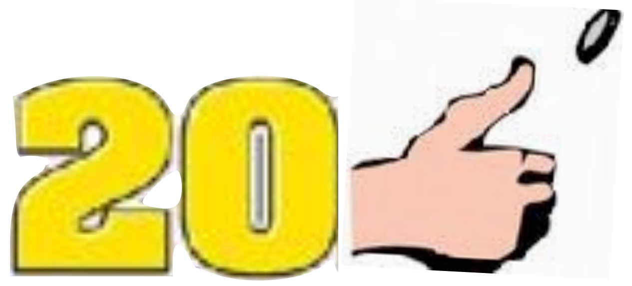
H_0 (x = 5)



Two-Sided Alternative

$H_0 : p_h = 0.5$ — Unbiased

$H_A : p_h \neq 0.5$ — Two-sided

Data 

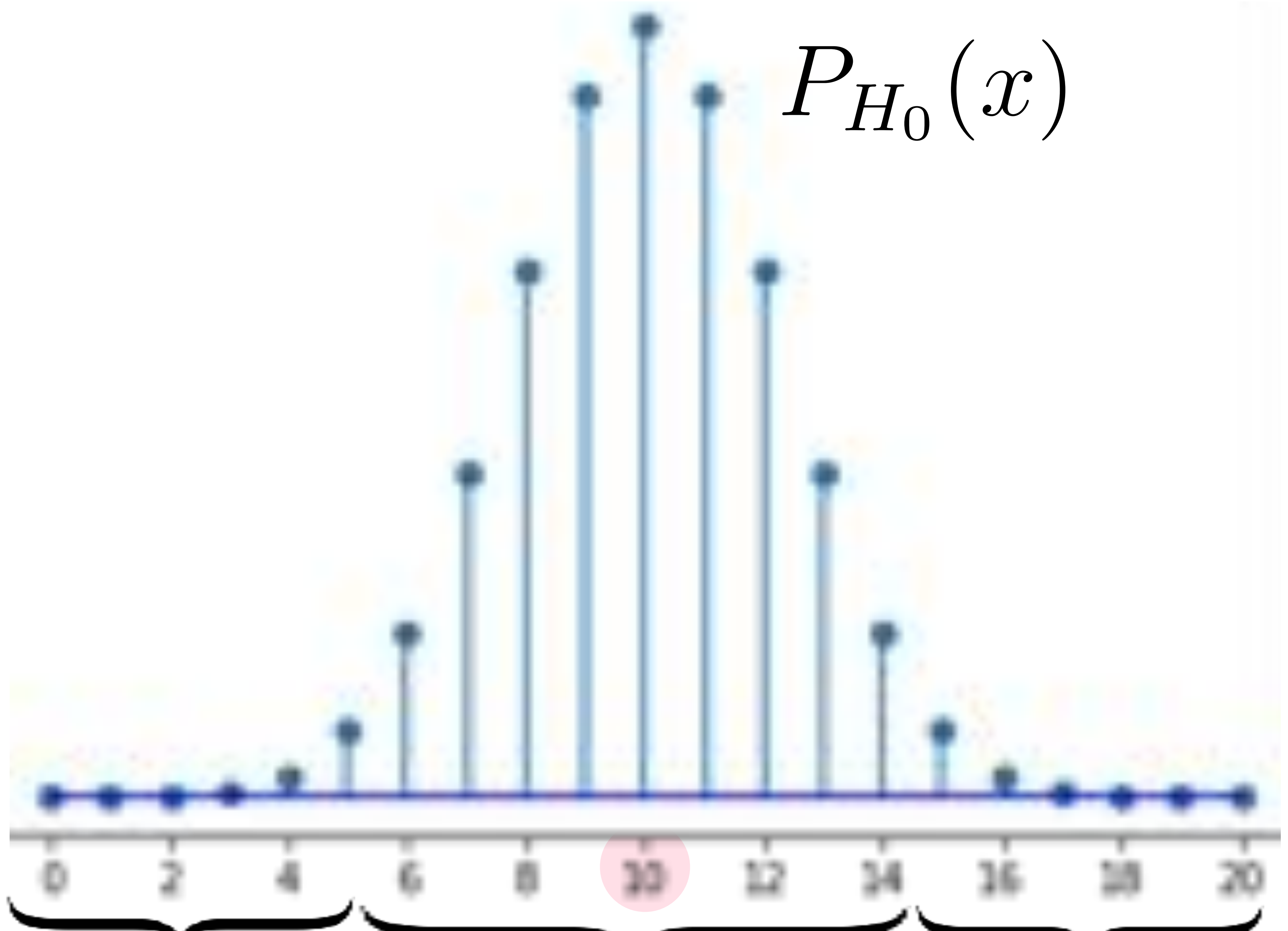
X : # 

H_0 Mean $\mu_X = 10$

Intuition

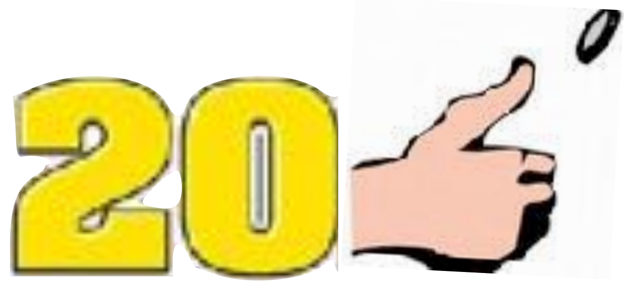
Under
- H_0: X close to 10
- H_A: X far from 10

$|X - 10|$
- small: Retain H_0
- large: Accept H_A



Critical Value

$H_0 : p_h = 0.5$ $H_A : p_h \neq 0.5$ Data



X : #

Significance level α Upper bound on type-I error

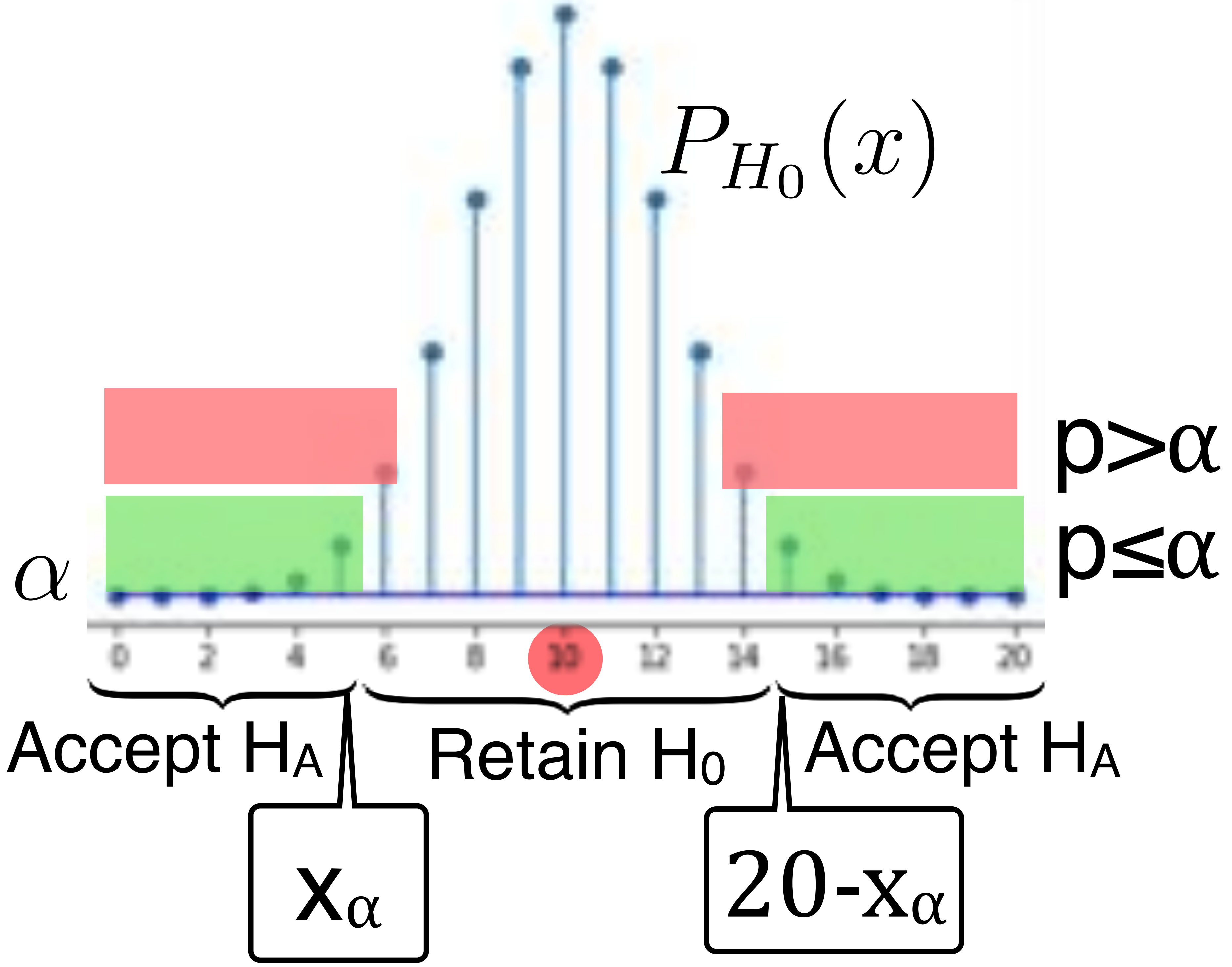
H_0 Mean $\mu_X = 10$

Critical value x_α x closest to 10 s.t. $P_{H_0}(|X-10| \geq |x-10|) \leq \alpha$

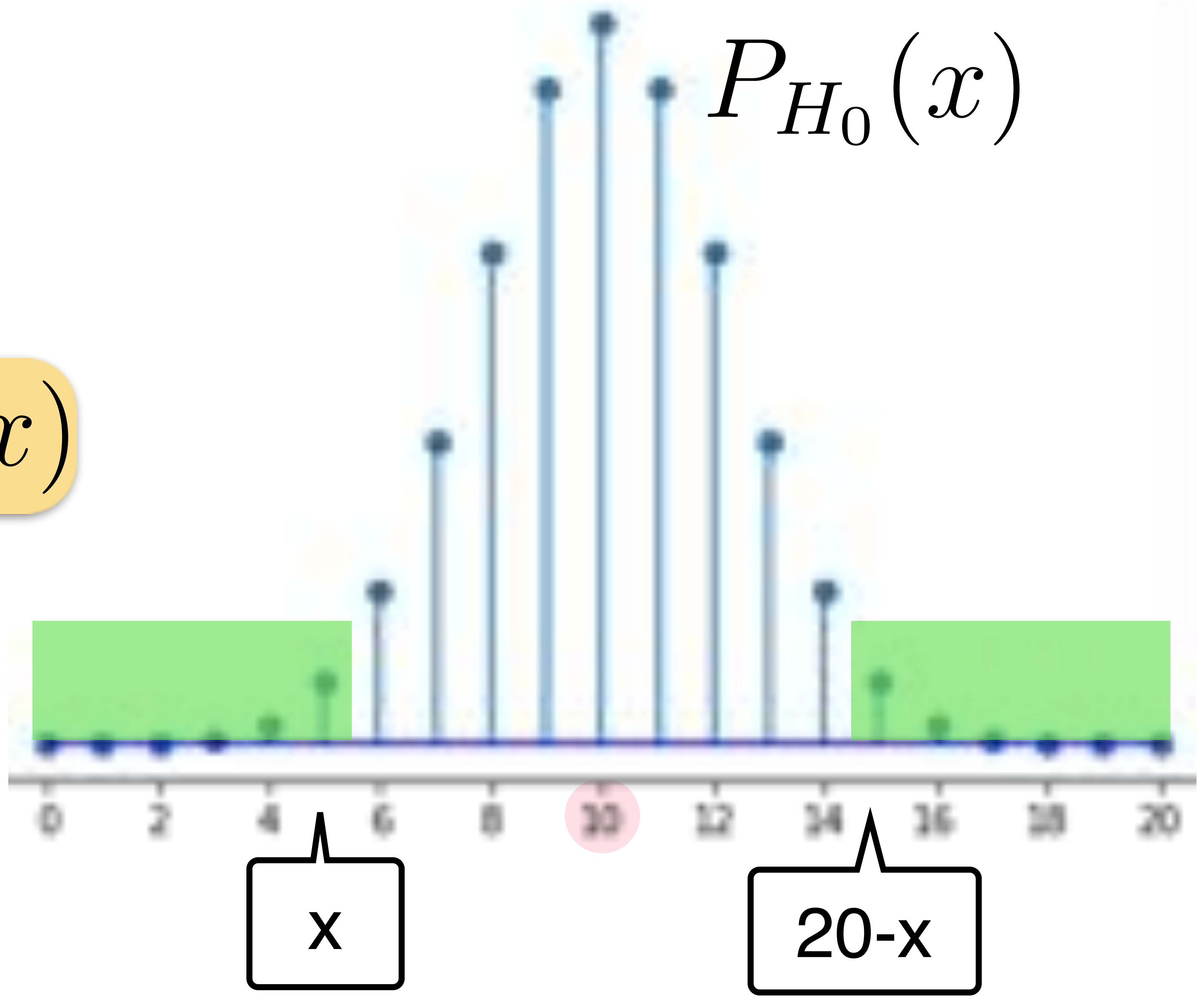
$$|X-10| \begin{cases} \geq |x_\alpha - 10| & \text{Accept } H_A \\ < |x_\alpha - 10| & \text{Retain } H_0 \end{cases}$$

P_{H_0}(type-I error) $= P_{H_0}(|X-10| \geq |x_\alpha - 10|) \leq \alpha$

x_α closest to 10 minimizes type-II error



p Values



p value of x

$P_{H_0}(X \text{ is at least as far from 10 as } x)$

$P_{H_0}(|X - 10| \geq |x - 10|)$

Low p value

x far from mean

Low H_0 prob of outcome x or further towards H_A

x less likely to be generated under H_0

High p value

x far from mean

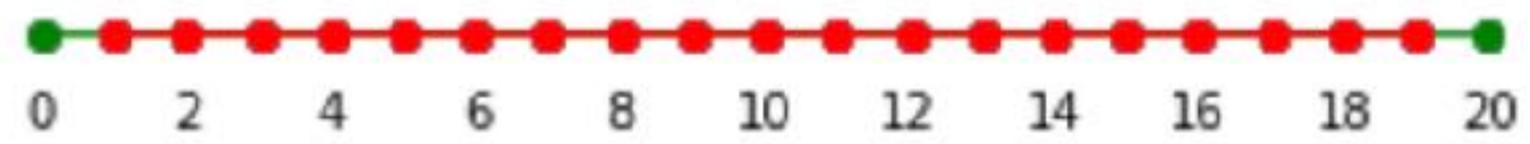
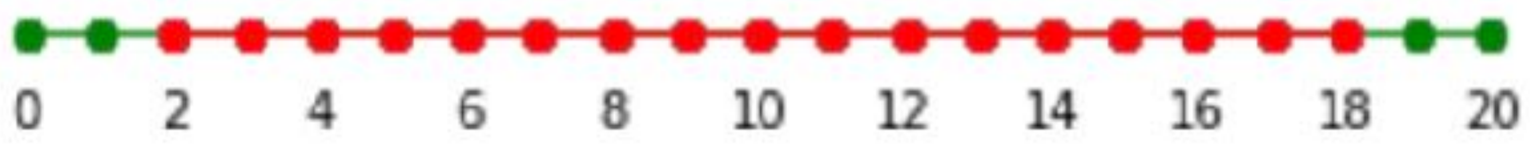
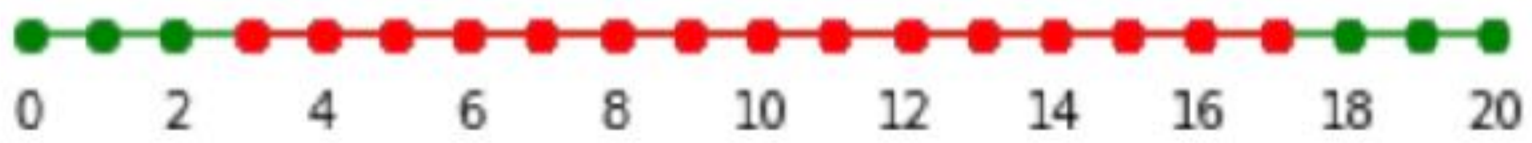
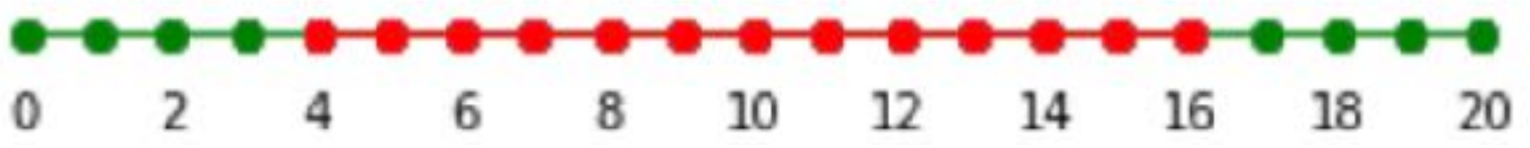
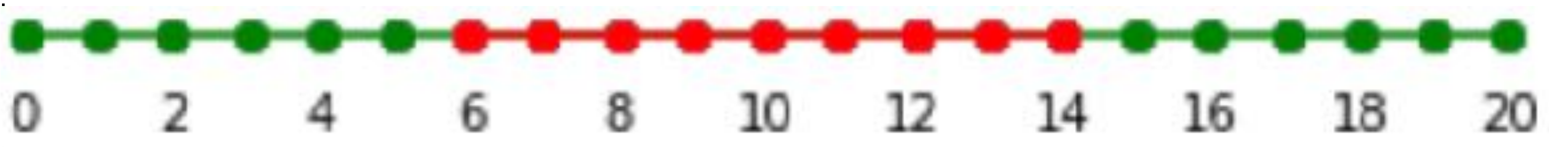
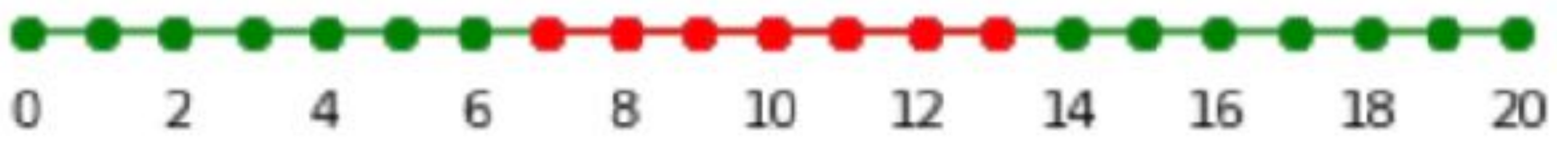
High H_0 prob of outcome x or further towards H_A

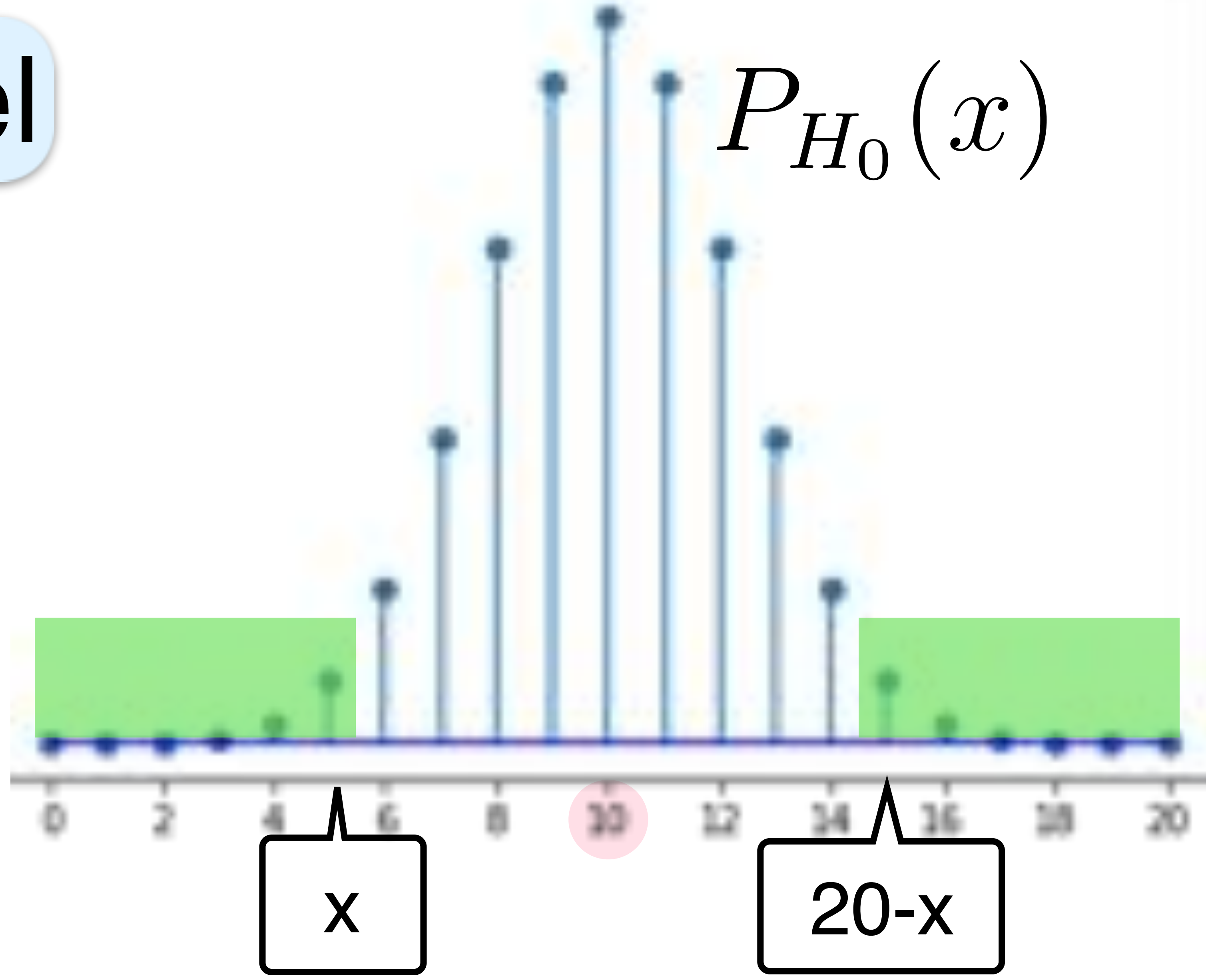
x more likely to be generated under H_0

p Values → Hypotheses

p value of x $P_{H_0}(X \text{ is at least as far from } 10 \text{ as } x)$ $P_{H_0}(|X-10| \geq |x-10|)$

5% Significance level

x	|X-10| ≥ |x-10|	p value of x	
0,20		0.0002%	H_A
1,19		0.0040%	•
2,18		0.0402%	•
3,17		0.2577%	•
4,16		1.1818%	•
5,15		4.1389%	H_A
6,14		11.5318%	H_0



p value ≤ 5% Accept H_A

p value > 5% Retain H_0

General p value

p value of statistic t of T

P_{H_0}(T is t or further towards H_A)

Significance level

α

p value

$\leq \alpha$

Accept H_A

$> \alpha$

Retain H_0

Hypothesis Testing - p Values

More systematic & precise

Significance level

Critical values

p values

One- two- sided alternatives