Knowing when and where to get help

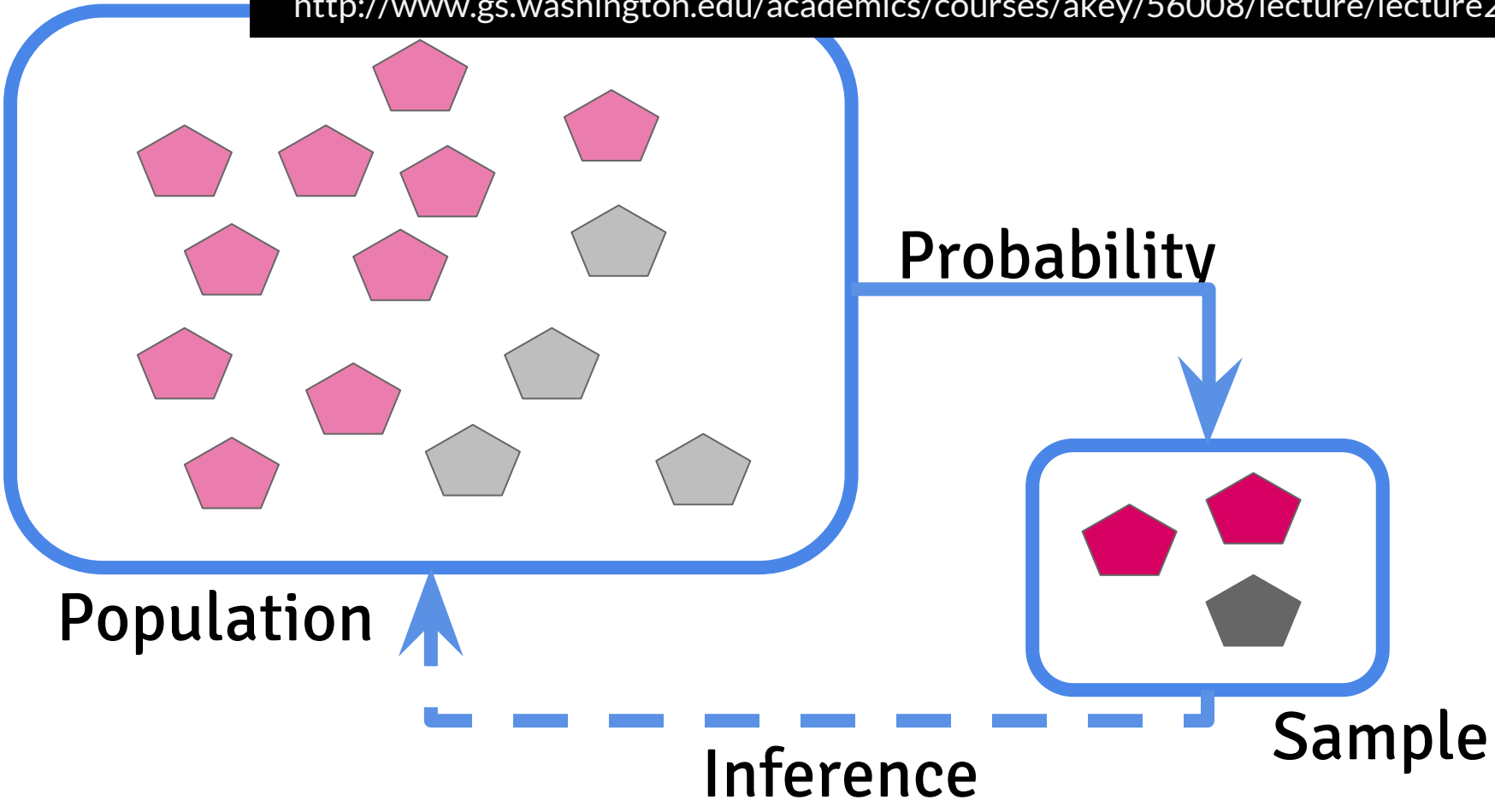
Jeff Leek

@jtleek

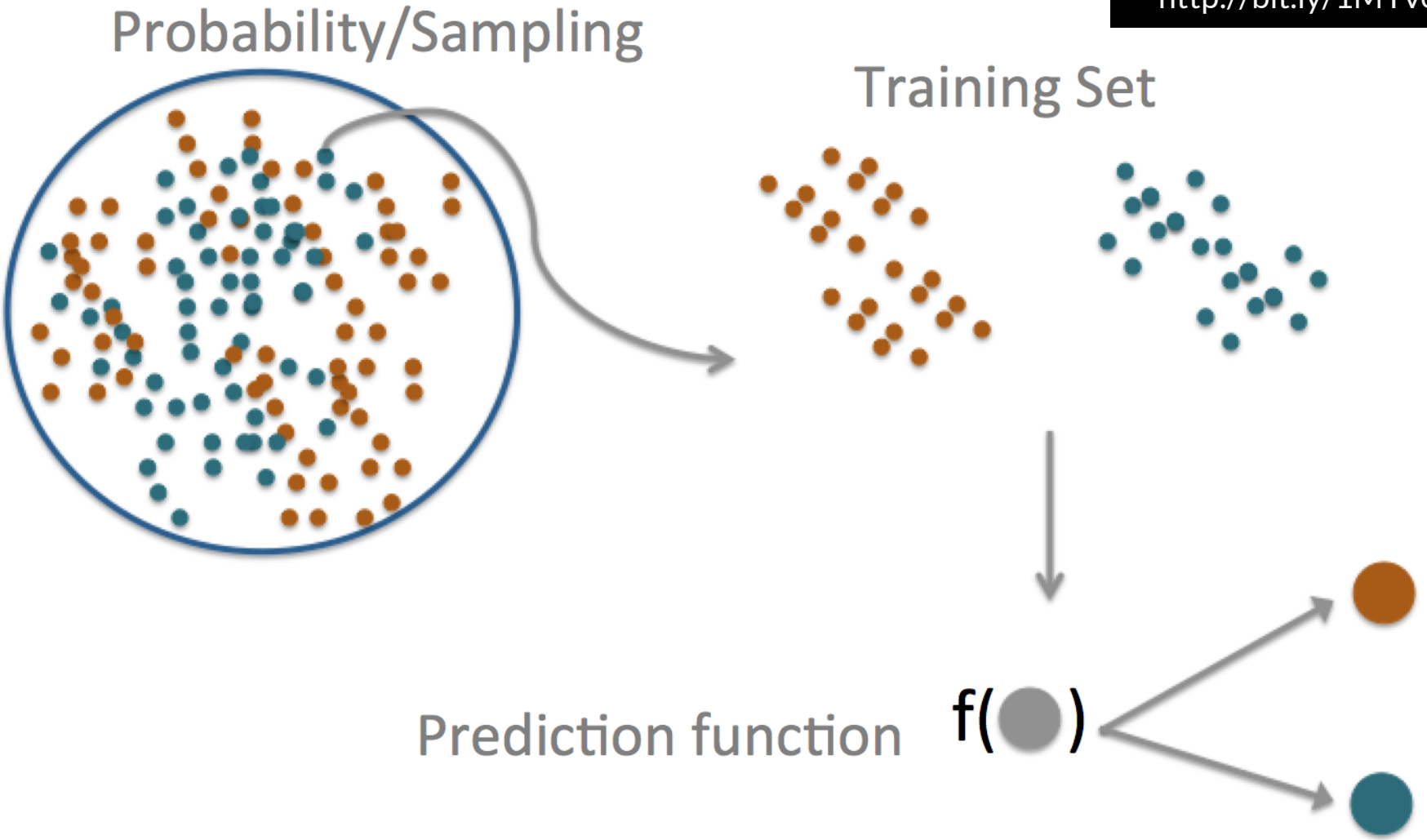
www.jtleek.com

Inference vs. prediction

http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture2.pdf

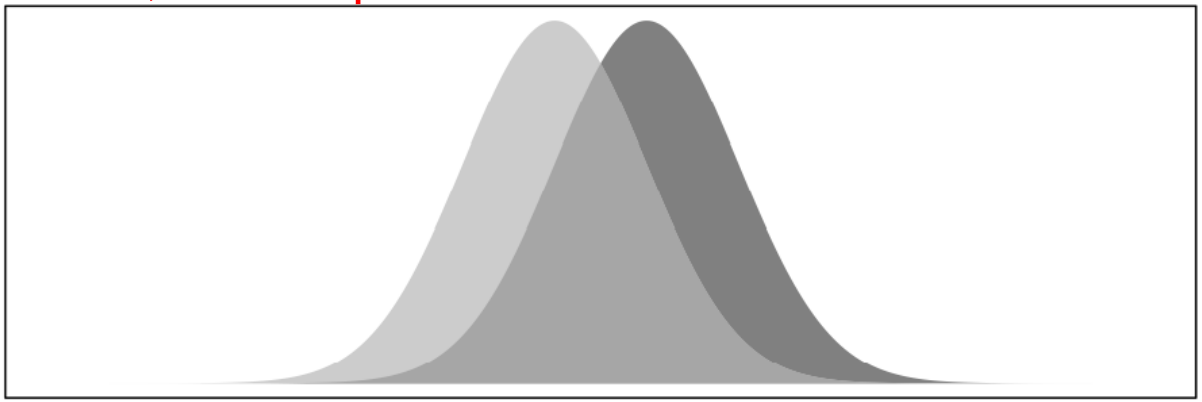


http://bit.ly/1MYve52

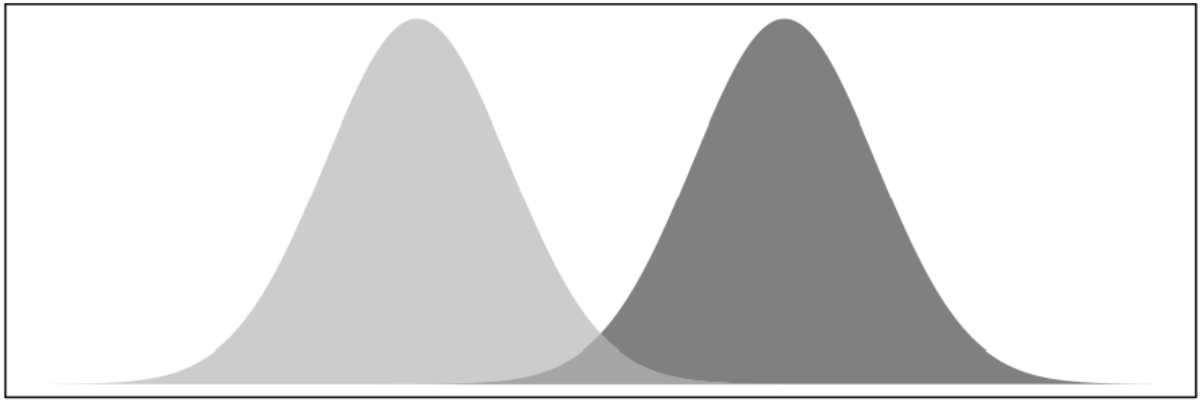


Slide courtesy: Ingo Ruczinski

Definitely different, but not predictive

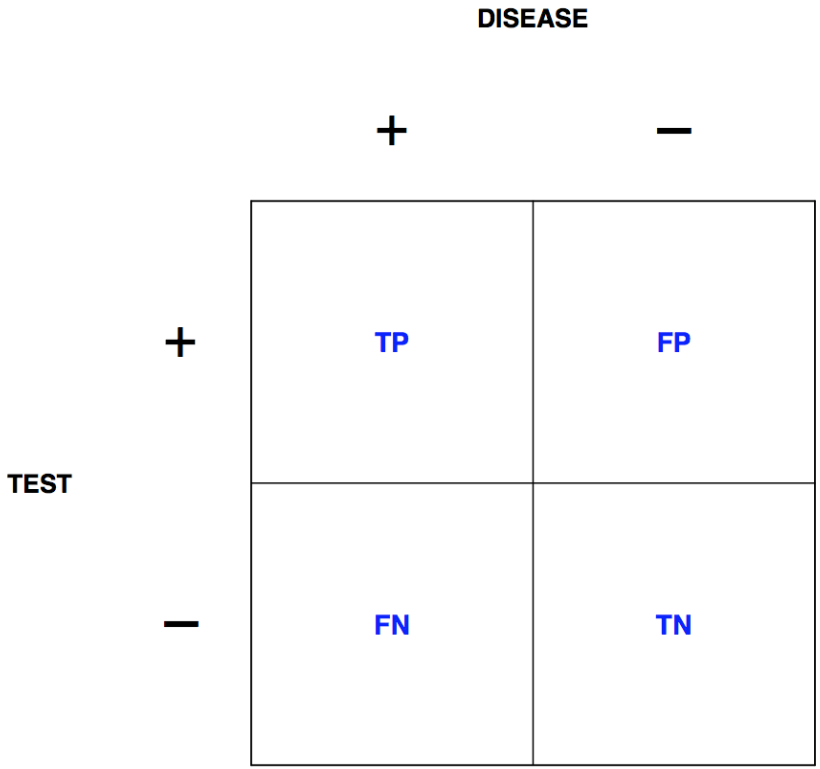


Different and predictive

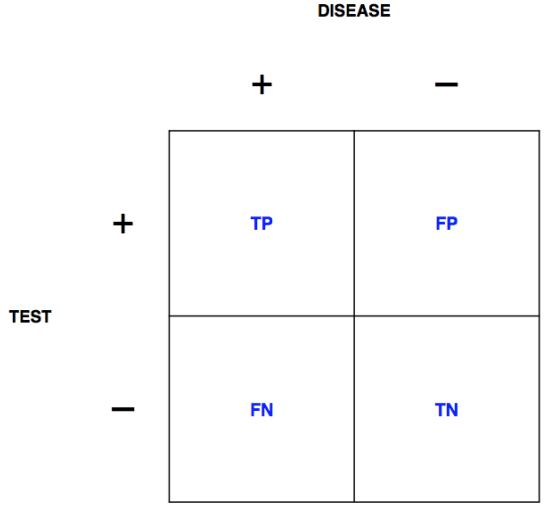


Key quantities

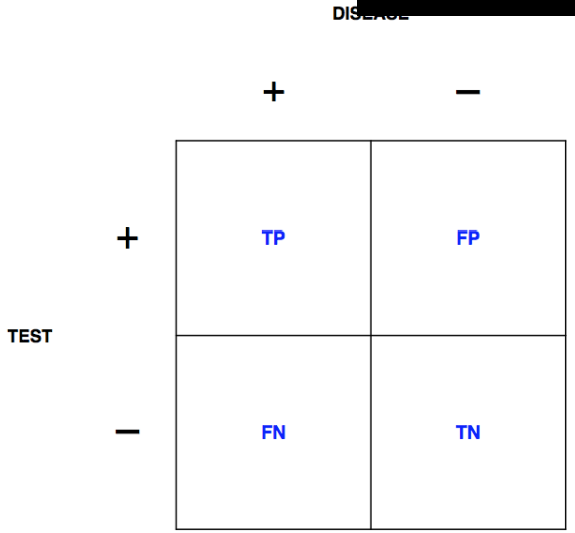
Slide courtesy: Ingo Ruczinski



Slide courtesy: Ingo Ruczinski



Sensitivity	→	Pr (positive test | disease)
Specificity	→	Pr (negative test | no disease)
Positive Predictive Value	→	Pr (disease | positive test)
Negative Predictive Value	→	Pr (no disease | negative test)
Accuracy	→	Pr (correct outcome)



Sensitivity → TP / (TP+FN)

Specificity → TN / (FP+TN)

Positive Predictive Value → TP / (TP+FP)

Negative Predictive Value → TN / (FN+TN)

Accuracy → (TP+TN) / (TP+FP+FN+TN)

An example

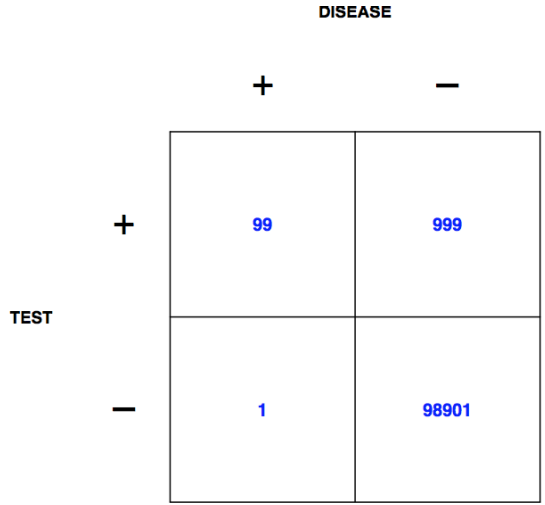
Assume that some disease has a 0.1% prevalence in the population. Assume we have a test kit for that disease that works with 99% sensitivity and 99% specificity. What is the probability of a person having the disease given the test result is positive, if we randomly select a subject from

- the general population?
- a high risk sub-population with 10% disease prevalence?

Slide courtesy: Ingo Ruczinski

TEST	DISEASE +	DISEASE −
+	99	999
−	1	98901

Slide courtesy: Ingo Ruczinski



Sensitivity	$\rightarrow$	99 / (99+1) = 99%
Specificity	$\rightarrow$	98901 / (999+98901) = 99%
Positive Predictive Value	$\rightarrow$	99 / (99+999) $\approx$ 9%
Negative Predictive Value	$\rightarrow$	98901 / (1+98901) $>$ 99.9%
Accuracy	$\rightarrow$	(99+98901) / 100000 = 99%

HEALTH

Vast Study Casts Doubts on Value of Mammograms

By GINA KOLATA FEB. 11, 2014

At risk subpopulation

TEST	DISEASE +	DISEASE −
+	9900	900
−	100	89100

Slide courtesy: Ingo Ruczinski

		DISEASE +	DISEASE −
TEST	+	9900	900
	−	100	89100

Sensitivity → 9900 / (9900+100) = 99%

Specificity → 89100 / (900+89100) = 99%

Positive Predictive Value → 9900 / (9900+900) ≈ 92%

Negative Predictive Value → 89100 / (100+89100) ≈ 99.9%

Accuracy → (9900+89100) / 100000 = 99%

Notes and further reading

- Prediction is a whole class (no joke): https://www.coursera.org/course/predmachlearn
- Prediction with genomics underlies precision medicine
- So far this has been a major challenge