

see syllabus

Case Study Credit Card Screening

good credit card declared bad: "false positive"
 bad credit card declared good: "false negative"

2 kinds of info:

(truth) \rightarrow B = (card really is good)
 (what system says) \rightarrow G = (card is really good)
 $\rightarrow \oplus$ = (system says bad)
 $\rightarrow \ominus$ = (system says good)

1% = prevalence = $P(B)$
 97% = $P(\ominus | G)$ "specificity"

1st way

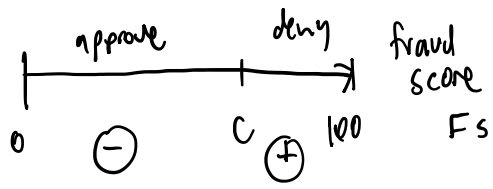
(2x2 contingency table) cross-tabulation of truth vs system

	truth B	truth G	
system says \oplus	98	297	395
system says \ominus	2	9,603	9,605
	100	9,900	10,000 ← transactions

sensitivity $98\% = P(\oplus | B)$ "sensitivity"
 specificity

we want $P(B | \oplus) = \frac{98}{395} \approx 25\%$

(false positive rate) $= P(G | \oplus) = 75\%$



(false negative rate) $= P(B | \ominus) = \frac{2}{9,605} \approx 0.02\%$

is the card bad? Fact (Statistical Inference)
 should we deny the transaction? Choice (Bayesian decision theory)
 ↑
 system

	truth B	truth G	
system \oplus	ok		agents: {customer, merchant, bank}
system \ominus		ok	

bank suffers small loss (bad) but not really bad

(card bad but system says good) (bad) much worse to bank

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n)$$

Simulation-based approximate probabilities

- 1750 Buffon needle
- 1908 W. Gosset
- 1941 Metropolis & Ulam (Monte Carlo Sampling)
- John Von Neuman algorithms for pseudo-random # generation

Fact: All probabilities are conditional, conditional on assumptions

$P(H) = \text{undefined}$ $P(H | \text{fair coin tossing}) = \frac{1}{2}$

chain rule for $A = \text{and}$

$$P(A_1 \cap \dots \cap A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \dots P(A_n | A_1 \cap \dots \cap A_{n-1})$$

hard

$$P(A) = P(A \text{ and } B_1 \text{ OR } A \text{ and } B_2 \text{ OR } \dots \text{ OR } A \text{ and } B_k)$$

simple addition rule for OR

$$= P(A \text{ and } B_1) + P(A \text{ and } B_2) + \dots + P(A \text{ and } B_k)$$

$$= P(B_1) \cdot P(A | B_1) + P(B_2) \cdot P(A | B_2) + \dots + P(B_k) \cdot P(A | B_k) \quad (LTP)$$