(see syllabus)

(see syllabus)

AMS 131
18 Apr A

①

case study

credit card screening

2 kinds of info:

(truth)

B = (card really is bad)

G = (card really is good)

(system says)

⊕ = (system says bad)

⊖ = (system says good)

1% = (prevalence) = P(B)

97% = P(⊖ | G) (specificity)

98% = P(⊕ | B) (sensitivity)

specificity

~~sensitivity~~

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

		truth B	truth G	
system says	⊕	98	297	395
	⊖	2	9,603	9,605
prevalence		100	9,900	10,000

sensitivity

← transactions

2

we want $P(B \mid \oplus) = \frac{98}{395} \doteq 25\%$ (!)

(false positive rate) $= P(G \mid \oplus) \doteq 75\%$ (!)

(false negative rate) $= P(B \mid \ominus) = \frac{2}{9605}$

$\doteq .00021$

$= 0.02\%$

---

approve | deny | fraud score

0 ——— c ——— 100 FS

is the card bad?

$\boxed{\text{fact}}$ (statistical inference)

system

should we deny the transaction? ← $\boxed{\text{choice}}$

Bayesian decision theory

	truth: B	G
system says +	ok	
−		ok

agents: 3
- customer
- merchant
- bank (building system)

bank suffers small loss (bad) but not really bad

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

Simulation-based approximate probabilities

- ~(1750) Buffon needle
- (1908) W. Gosset (Monte Carlo sampling)
- (1941) — Metropolis & Ulam
  — John von Neumann
  algorithms for pseudo-random # generation

fact: All probabilities are conditional; conditional on assumptions

$P(H) =$ undefined

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

hard

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

(no-overlap) simple addition rule for or

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

$= \underbrace{P(B_1)}_{\text{easy}} \cdot \underbrace{P(A \mid B_1)}_{\text{easy}} + P(B_2) \cdot P(A \mid B_2) + \dots + P(B_k)\, P(A \mid B_k)$

$$P(A) = \sum_{j=1}^{k} P(B_j)\, P(A \mid B_j) \quad \text{(LTP)}$$