4/18/19 Lecture 6

Credit Card Screening Case Study

false positive: when a good credit card is declared bad

false negative: when a bad credit card is declared good

97% of the test purchases were correctly labeled good

98% of the test purchases were correctly labeled bad

1% of all attempted purchases are with bad credit cards

Events: B: card really is bad] Truth
G: card really is good]

S+: System says bad] System
S-: System says good]

0.01 = prevalence = $P(B)$

0.97 = specificity = $P(S- | G)$

0.98 = sensitivity = $P(S+ | B)$

2×2 Contingency table aka cross-tabulation of truth vs the system

Truth

System	B	G	total
S+	98 (5)	297 (4)	395 (7)
S-	2 (6)	9,603 (3)	9,605 (8)
total	100 (1)	9,900 (2)	10,000 (transactions)

$$\left.\begin{array}{ll} (1) & 0.01 \times 10{,}000 = 100 \\ (2) & 10{,}000 - 100 = 9{,}900 \end{array}\right\} \text{prevalence}$$

$$\left.\begin{array}{ll} (3) & 0.97 \times 9{,}900 = 9{,}603 \\ (4) & 9{,}900 - 9{,}603 = 297 \end{array}\right\} \text{specificity}$$

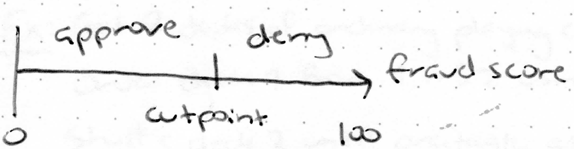
$$\left.\begin{array}{ll} (5) & 0.98 \times 100 = 98 \\ (6) & 100 - 98 = 2 \end{array}\right\} \text{sensitivity}$$

$$\left.\begin{array}{ll} (7) & 98 + 297 = 395 \\ (8) & 2 + 9{,}603 = 9{,}605 \end{array}\right\} \text{sums}$$

We want $P(B \mid S+) = \dfrac{98}{395} \approx 25\%$ (?!?!)

False positive rate $= P(G \mid S+) \approx 75\%$ (complement)

False negative rate $= P(B \mid S-) = \dfrac{2}{9605} \approx 0.02\%$



Is the card bad?
Based on fact (statistical inference)

Should the system deny the transaction?
Based on choice (Bayesian decision theory) ← More than just the truth

Truth

		B	G
System	S+	ok	mistake
	S−	mistake	ok

agents:
- • customer
- • merchant
- ★ bank

The bank builds the system. (with themselves in mind).

G and S+ : bank suffers a small loss (bad but not really bad)
B and S− : much worse to the bank

the mistakes

The only way to make the system better is to have a better fraud score.

Bank prefers the cardholder to suffer than them. Sorry.

Generalization: For any events $A_1, \dots, A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j)$$

$$+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \dots$$

$$+ (-1)^{n+1} P(A_1 \cap \dots \cap A_n)$$

<u>Ex:</u> Get 2 decks of ordinary playing cards.

Order deck 1 from 1 to 52 using any sequence you like.

Shuffle deck 2 until practically all 52! orderings are equally likely.

Now turn over the first card of each deck. Do they match? Continue through all 52 cards

P(at least one match) = ? let $n = 52$

Let A_i = a match occurs on card i

We want $P\left(\bigcup_{i=1}^{n} A_i\right)$ which can be computed with the complicated formula above.

Problems can be answered with math or simulation

Ex. of simulation: Buffon's needle problem - approx. π

Back to Example: Calculus Result: $\lim_{n \to \infty} \sum_{i=1}^{n} \frac{(-1)^{i+1}}{i!} = 1 - \frac{1}{e} = 0.63$

This sum approaches its limit quickly; already with $n=7$ you have the first 4 significant figures: 0.6321

Conditional Probability

Textbook (ch. 2) Note that Kolmogorov's probability axioms defined the function $P_K(A)$, where A is a set in the collection $\mathcal{C}$ of subsets of the sample space S in which nothing weird can occur

In other words, $P_K(A)$ is a function of a single argument A.

To include the extremely useful idea of conditional probability in his setup, Kolmogorov has to define it using P_K.

Def: Given any two events A, B in $\mathcal{C}$, the conditional probability of A given B is

$$P(A \mid B) = \begin{cases} \frac{P(A \cap B)}{P(B)} & \text{if } P(B) > 0 \\ \text{undefined} & \text{if } P(B) = 0 \end{cases}$$

There are other foundational theories of probability - one by Bruno de Finetti and another by Richard T. Cox and Edwin T. Jaynes - in which the probability function $P_{deF}(A, B)$ or $P_{CJ}(A \mid B)$ has 2 inputs, not 1, so that conditional probability is the primitive concept, not unconditional probability or with Kolmogorov's $P_K(A)$

Fact: All probabilities are conditional — conditional on assumptions

Ex: Tossed a penny. What's the prob. it turns out heads?

It's not $\frac{1}{2}$. It's $\frac{1}{2}$ conditional on the assumption that the coin tossing is fair. So $P(H)$ = undefined.

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

All probabilities are conditional on 3 things:

1) background assumptions (A)

2) background information (I)

3) background judgements (J)

<u>Ex</u>: T-S) We actually computed not

$P(\text{at least } 1 \text{ T-S baby})$ but

$P(\text{at least } 1 \text{ T-S baby} \mid \text{family of 5 \& mother \& father both carriers})$

This impulse to be explicit about your AIJ is <u>Bayesian</u>; Kolmogorov worked in the <u>frequentist</u> paradigm

In this course, focusing on $P_K(B)$, we need to remember that it should really be $P_K(B \mid AIJ)$

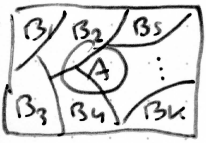
<u>Consequences of the Conditional Probability Definition (Theorems)</u>

① A, B events in C: if $P(B) > 0$ then $P(A \cap B) = P(B)P(A \mid B)$

and if $P(A) > 0$ then $P(A \cap B) = P(A)P(B \mid A)$

② Direct Generalization: if $A_1, \ldots, A_n$ are events with $P(A_1 \cap \ldots \cap A_{n-1}) > 0$ then <u>Chain Rule for n = AND</u>

$$P(A_1 \cap \ldots \cap A_n) = P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1 \cap A_2) \ldots P(A_n \mid A_1 \cap \ldots \cap \ldots A_{n-1})$$



Hard

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

Simple (no overlap) addition rule for OR

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

$= P(B_1) \cdot P(A \mid B_1) + P(B_2) \cdot P(A \mid B_2) + \ldots$

Easy

$+ P(B_k) \cdot P(A \mid B_k)$

$$P(A) = \sum_{j=1}^{k} P(B_j)\,P(A \mid B_j)$$ Law of total probability (LTP)

<u>When is LTP useful?</u>

You're trying to compute $P(A)$ and you find it hard to compute directly. If you can find some aspect B of the world satisfying 2 properties —

① B defines a partition $\{B_1, \ldots, B_k\}$ of S with known $P(B_j)$

② A depends on B in such a way that the conditional probabilities $P(A \mid B_j)$ are easier to compute than $P(A)$ itself — then you can work out $P(A \cap B_j)$

$P(A)$ indirectly: $P(A) = \sum_{j=1}^{k} \underbrace{P(B_j)\,P(A \mid B_j)}_{P(A \cap B_j)}$ (Bayesian mixture modeling)