T-61.5040 Oppivat mallit ja menetelmät
T-61.5040 Learning Models and Methods
Pajunen, Viitaniemi

Solutions to exercise 3, 2.2.2007

Problem 1.

Denote the propositions as S = innocent, T = fingerprints are identical, M = guilty. We wish to compute $p(S|T)$, which is according to Bayes theorem

$$p(S|T) = \frac{p(T|S)p(S)}{p(T)}$$

The terms in the above formula are:
$p(S) = 1 - 10^{-7} = 10^{-7}(10^7 - 1)$, as there is one burglar among 10 million inhabitants
$p(T) = p(T|S)p(S) + p(T|M)p(M)$ is the total probability that the fingerprints are identical. Here
$p(T|S) = \frac{9}{10^7-1}$, as there are $10 - 1 = 9$ people with identical fingerprints as the burglar among the $10^7 - 1$ innocent citizens.
$p(T|M) = 1$, as the fingerprints of the burglar are of course identical with themselves,
$p(M) = 10^{-7}$, as we assume there is exactly one burglar.
We get

$$p(S|T) = \frac{\frac{9}{10^7-1}10^{-7}(10^7-1)}{\frac{9}{10^7-1}10^{-7}(10^7-1) + 10^{-7}} = \frac{9 \cdot 10^{-7}}{10 \cdot 10^{-7}} = 0.9$$

If a suspect is sentenced of a burglary based only on identical fingerprints, the probability of innocence is around 90 %!.

Example of wrong intuition: since fingerprint recognition is very reliable, the probability that you are innocent is very small.

This is wrong because we are not trying to find out the probability $P(T|S)$, that is "the probability that your fingerprint matches *if you are innocent*". It would only tell that for a randomly picked innocent citizen it is unlikely that the fingerprints match. You were not picked randomly, you were among the people with matching fingerprints.

This problem illustrates that even in simple problem settings intuitition can lead very far from the correct answer. One must be very careful in formulating the problem in terms of conditional probabilities, and computing the right probability.

Problem 2

Yes, you should change your choice, because that increases your probability of having the prize:
Denote the correct door by A. Then the prior probabilities are $p(A = 1) = p(A = 2) = p(A = 3) = 1/3$. We wish to know the posterior of A given the observation.

Choose door number 1. Then assume that door 2 is opened and there is no prize behind that. Now we should compute $p(A|D)$ where D means "door number 2 was opened". We know that $p(A = 2|D) = 0$.

By Bayes' rule, $p(A|D) = p(D|A)p(A)/p(D)$. We are interested in A so it is enough that $p(A|D) \propto p(D|A)p(A)$, as $p(D)$ is merely a scaling term.

$p(D|A = 1) = 1/2$, because if the prize is behind the door that you had already chosen, then either of the doors 2 and 3 is opened with equal probability.

$p(D|A = 3) = 1$, because if you had chosen door 1 and the prize is behind door 3, then the only possibility is to open door 2.

$p(D|A = 2) = 0$, because door 2 is not opened if the prize is behind it.

Thus $p(A = 1|D) \propto p(D|A = 1)p(A = 1) = \frac{1}{2} * \frac{1}{3}$
and $p(A = 3|D) \propto p(D|A = 3)p(A = 3) = 1 * \frac{1}{3}$.
You increase your probability of getting the prize, if you change to door 3!

Problem 3.

i) Suppose A_1 and A_2 equivalent events. Then Cox's axiom requires that $p(A_1) = p(A_2)$. Violating the axiom requires $q(A_1) \neq q(A_2)$. Let $q(A_1) > q(A_2)$ (otherwise swap the roles of the variables in the following). If you buy T_1 and sell T_2 at your the limit price, you have paid $q(A_1) - q(A_2) > 0$ EUR. The Bayesian will agree to the trade since the net payment he receives is > 0 EUR but the sum of the winning probabilities on the tickets he gives away is 0. But A_1 and A_2 are equivalent, so either both tickets win or both tickets lose. In both cases your winnings are zero, so you have paid a positive amount of money to obtain nothing.

ii) Suppose $q(A_1) < 1 (= p(A_1))$. Then selling T_1 at the limit price gives you $q(A_1)$ EUR but when A_1 happens, you have lost 1 EUR making a net loss of $1 - q(A_1)$ EUR. If $q(A_1) > 1$, then you are willing to pay more than 1 EUR for T_1, which pays you only 1 EUR.

iii) The sum rule says that for two mutually exclusive but exhaustive events A_1 and A_2, $p(A_1) + p(A_2) = 1$. Violating the assumption thus requires that $q(A_1) + q(A_2) = S \neq 1$. Let $S < 1$. Then the Bayesian is willing to buy tickets T_1 and T_2 from you since your total price S is less than the winning probability $p(A_1) + p(A_2) = 1$. After observing the outcomes you pay him 1 EUR with certainty. You thus have lost $1 - S$ EUR. On the other hand, if $S > 1$ the Bayesian is willing to sell you tickets for your price. Once again, you lose $S - 1$ EUR with certainty.

Comments: Just as in the case of the sum rule, it would be possible to show that violating the product rule $p(AB|C) = p(A|BC)p(B|C)$ leads to a Dutch Book. However, the related calculation is more complicated and is skipped here. Also, violating $p(A) \in \mathbb{R}$ or $p(A) \geq 0$ leads to a Dutch Book.

This problem illustrates the Dutch Book Theorem, which formally says that a Dutch

Book can be constructed if and only if q is not a probability. We have now shown that if q lacks any of the basic properties of a probability measure then a Dutch Book follows. Proving that the possibility of a Dutch Book leads into a non-probability measure is a more difficult task.

Problem 4.

i) The iteration formula $\mathrm{Var}(x) = \mathrm{E}(\mathrm{Var}(x|y)) + \mathrm{Var}(\mathrm{E}(x|y))$ says something about distributions $p(x)$ and $p(x|y)$. Substitute $x = \theta$ and $y = D$ to obtain

$$\mathrm{Var}(\theta) = \mathrm{E}(\mathrm{Var}(\theta|D)) + \mathrm{Var}(\mathrm{E}(\theta|D)).$$

The left-hand side is the variance of the prior distribution. The right-hand side has two terms where both terms are nonnegative. The expectation $\mathrm{E}(\mathrm{Var}(\theta|D))$ is an expectation of the posterior variance taken over all data sets using the predictive distribution $p(D)$. This tells us the average posterior variance. Since the second term is nonnegative, we conclude that posterior variance on average cannot be larger than prior variance. Therefore on average Bayesian Inference is reducing uncertainty in θ.

ii) The number of terms in the sum is random: this seems difficult, but can easily be handled using the iteration formulas demonstrating their usefulness.

Recall that if a random variable $X \sim \mathrm{Poisson}(\lambda)$, we have $\mathrm{E}(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$. If a random variable $Y \sim \mathrm{Exp}(\mu)$, we have $\mathrm{E}(Y) = 1/\mu$ and $\mathrm{Var}(Y) = 1/\mu^2$.

The mean of s is

$$\mathrm{E}(s) = \mathrm{E}(\mathrm{E}(s|N)).$$

Conditional to N, the sum s has a known number of terms. Therefore $\mathrm{E}(s|N) = N\mathrm{E}(s_1) = N/\mu$. To find $\mathrm{E}(s)$, we compute

$$\mathrm{E}(s) = \mathrm{E}(N/\mu) = \lambda/\mu.$$

Then the variance:

$$\mathrm{Var}(s) = \mathrm{E}(\mathrm{Var}(s|N)) + \mathrm{Var}(\mathrm{E}(s|N)).$$

The last term is $\mathrm{Var}(N/\mu) = \mu^{-2}\mathrm{Var}(N) = \mu^{-2}\lambda$. In the first term, the variance of s given N is $N\mathrm{Var}(s_1) = N\mu^{-2}$. The mean of $\mathrm{Var}(s|N)$ is then $\mathrm{E}(N\mu^{-2}) = \mu^{-2}\lambda$. Adding them up we obtain

$$\mathrm{Var}(s) = 2\lambda\mu^{-2}.$$