Statistical Methods (Autumn 2017): Homework Problems I

October 2, 2017

1 Problem 1

1.1 Proof

Notation:

- $A \cup B$ is the union of A and B (A:n ja B:n yhdiste)
  - "A or B"
- $A \cap B$ is the intersection of A and B (A:n ja B:n leikkaus)
  - "A and B"
- $A \backslash B$ is the complement of A and B (A:n ja B:n joukkoerotus)
  - "A minus B"

The Kolmogorov axioms are (for a sample space S):

1. For every subset A in S, $P(A) \geq 0$
2. If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$
3. $P(S) = 1$

Show $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Let's split it up as a union of three disjoint sets:

$$P(A \cup B) = P((A \backslash (A \cap B)) \cup (B \backslash (A \cap B)) \cup (A \cap B))$$

Now let's use Kolmogorov's 2nd axiom:

$$= P(A \backslash (A \cap B)) + P(B \backslash (A \cap B)) + P(A \cap B)$$

$$= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B)$$

$$= P(A) + P(B) - P(A \cap B)_{\square}$$

1.2 Bayes' theorem

Abbreviations used:

- W Windows
- L Linux
- M MacOS
- V Infected by virus

Operating system market shares given as:

$$P(W) = 0.55$$

$$P(M) = 0.20$$

$$P(L) = 0.25$$

Probability of virus given a certain operating system:

$$P(V|W) = 0.08$$

$$P(V|M) = 0.03$$

$$P(V|L) = 0.02$$

Bayes' theorem states:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

The probability of a random infected user being a Windows user is $P(W|V)$. Using Bayes' theorem:

$$P(W|V) = \frac{P(V|W)P(W)}{P(V)}$$

The probability of a virus in general (for a user of any of the three operating systems) $P(V) = P(V|M) \cdot P(M) + P(V|W) \cdot P(W) + P(V|L) \cdot P(L) = 0.03 \cdot 0.2 + 0.08 \cdot 0.55 + 0.02 \cdot 0.25 = 0.055$.

$$P(W|V) = \frac{0.08 \cdot 0.55}{0.055} = \mathbf{0.8} = \mathbf{80}\,\%$$

Probability of an infected user not being a Mac user is given by:

$$P(\overline{M}|V) = 1 - P(M|V) = 1 - \frac{P(V|M) \cdot P(M)}{P(V)} = 1 - \frac{0.03 \cdot 0.2}{0.055} \approxeq \mathbf{0.8909} = \mathbf{89.1}\,\%$$

2 Problem 2

The distance between the centers of two strips is d. Particles enter perpindicularly according to a uniform distribution.

The following formulae apply for a uniform distribution:

$$f(x,\ \alpha,\ \beta) = \begin{cases} \frac{1}{\beta - \alpha} & \alpha \leq x \leq \beta \\ 0 & otherwise \end{cases}$$

$$\sigma^2 = V[x] = \int_\alpha^\beta [x - \frac{1}{2}(\alpha + \beta)]^2 \frac{1}{\beta - \alpha} = \frac{1}{12}(\beta - \alpha)^2$$

Let's study two adjacent strips n and $n-1$:

$$\sigma^2 = \frac{1}{12}(nd - (n-1) \cdot d)^2 = \frac{1}{12}d^2$$

Thus the standard deviation (or the uncertainty on the position given binary readout) is:

$$\sigma = \frac{d}{\sqrt{12}} \square$$