BAYESIAN PARADIGM

KALOYAN HARALAMPIEV

$$P(DH) = P(D|H).P(H)$$

$$P(HD) = P(H|D).P(D)$$

BAYES THEOREM

$$P(H|D) = \frac{P(D|H).P(H)}{P(D)}$$

$$P(H|DI) = \frac{P(D|HI).P(H|I)}{P(D|I)}$$

- Jacob Bernoulli (1654-1705)
  - Ars Conjectandi (1713)
- Thomas Bayes (1702-1761)
  - Essay towards Solving a Problem in the Doctrine of Chances (1763)

- Pierre Simon Laplace (1749-1827)

- Criticizm of Laplace:
  - John Venn (1834-1923)
  - Ronald Fisher (1890-1962)
  - Jerzy Neyman (1894-1981)
  - Karl Pearson (1857-1936)

- Harold Jeffreys (1891-1989)
  - Theory of Probability (1939)
- Richard Cox (1898-1991)
  - Probability, Frequency, and Reasonable Expectation (1946)

COX THEOREM

If:

1) The degrees of plausibility are represented by real numbers;
2) There is qualitative correspondence with common sense;
3) There is consistency.

Then follow:

- Product rule;
- Sum rule.

PRODUCT RULE

$$P(AB) = P(B|A).P(A) = P(A|B).P(B)$$

SUM RULE

$$P(A) + P(\bar{A}) = 1$$

COROLLARIES OF THE SUM RULE

- For 2 events A and B:
$P(A + B) = P(A) + P(B) – P(AB)$

- For 2 mutually exclusive events A and B:
$P(A + B) = P(A) + P(B)$

- For k mutually exclusive events $A_к$:
$P(\Sigma A_k) = \Sigma P(A_k)$

- For k mutually exclusive and exhaustive hypotheses $H_к$:
$P(\Sigma H_k) = \Sigma P(H_k) = 1$

MARGINAL PROBABILITY

$$\sum_k [P(D|H_k I).P(H_k|I)] = \sum_k P(DH_k|I) =$$

$$=P(D \textstyle\sum_k H_k|I) = P(D|I)$$

BAYES THEOREM AGAIN

$$P(H_i|DI) = \frac{P(D|H_iI).P(H_i|I)}{\sum_k [P(D|H_kI).P(H_k|I)]}$$

- Claude Shannon (1916-2001)
  - Mathematical Theory of Communication (1948)

INFORMATION ENTROPY

$$-K\sum_{k}[P(H_k|I).\,logP(H_k|I)]$$

Principle of maximum entropy

- Edwin Jaynes (1922-1998)
  - Propriety
  - Jaynes Consistency

ALGORITHM OF THE BAYESIAN STATISTICS

1. Hypotheses definition:

H_k: $\Theta = x_k$, $x_1 < x_2 < x_3 < \ldots$

2. Application of the principle of maximum entropy for the assigning of the prior probabilities $P(\Theta = x_k | I)$

ALGORITHM OF THE BAYESIAN STATISTICS

3. Choice of the sampling distribution

4. Calculation of the marginal probability

5. Application of the Bayes theorem for calculation of the posterior probabilities $P(\Theta = x_k | DI)$

ALGORITHM OF THE BAYESIAN STATISTICS

6. From the posterior probabilities which form probability density function we obtain cumulative probability density function

$$f(x_k) = P(\Theta = x_k | DI)$$

$$F(x_k) = P(\Theta \leq x_k | DI)$$

ALGORITHM OF THE BAYESIAN STATISTICS

7. Credible intervals and hypotheses testing

$$P(\Theta \leq a | DI) = F(a)$$

$$P(\Theta > b | DI) = 1 - F(b)$$

$$P(b < \Theta \leq a | DI) = F(a) - F(b)$$

ONE EXAMPLE

- Population with size N
- Sample with size n
- Random choice without replacement
- Nominal variable with m categories
- $\sum_{i=1}^{m} f_i = n$
- $\sum_{i=1}^{m} \hat{f}_i = N$

1. Hypotheses definition:

$$f_i \leq \hat{f}_i \leq f_i + (\mathrm{N} - \mathrm{n})$$

$$\frac{f_i}{N} \leq \frac{\hat{f}_i}{N} \leq \frac{f_i + (\mathrm{N} - \mathrm{n})}{N}$$

$$\frac{f_i}{N} \leq \pi_i \leq \frac{f_i + (\mathrm{N} - \mathrm{n})}{N}$$

$$\pi_{i,1} = \frac{f_i}{N}, \pi_{i,2} = \frac{f_i + 1}{N}, \pi_{i,3} = \frac{f_i + 2}{N}, \dots, \pi_{i,last} = \frac{f_i + (N - n)}{N}$$

2. Using of the principle of maximum entropy for the assigning of the prior probabilities

$$const$$

3. Choice of the sampling distribution

$$\frac{C_{\hat{f}_1}^{f_1} C_{\hat{f}_2}^{f_2} \dots C_{\hat{f}_m}^{f_m}}{C_N^n}$$

4. Calculation of the marginal probability

$$\sum\left(\frac{C_{\hat{f}_1}^{f_1} C_{\hat{f}_2}^{f_2} \ldots C_{\hat{f}_m}^{f_m}}{C_N^n} const\right) = \frac{\sum\left(C_{\hat{f}_1}^{f_1} C_{\hat{f}_2}^{f_2} \ldots C_{\hat{f}_m}^{f_m}\right)}{C_N^n} const =$$

$$= \frac{C_{N+m-1}^{N-n}}{C_N^n} const$$

5. Application of the Bayes theorem for calculation of the posterior probabilities

$$\frac{\frac{C_{\hat{f}_1}^{f_1} C_{\hat{f}_2}^{f_2} \dots C_{\hat{f}_m}^{f_m}}{C_N^n} const}{\frac{C_{N+m-1}^{N-n}}{C_N^n} const} = \frac{C_{\hat{f}_1}^{f_1} C_{\hat{f}_2}^{f_2} \dots C_{\hat{f}_m}^{f_m}}{C_{N+m-1}^{N-n}}$$

$$P(H_1|DI) = \frac{C_{\hat{f}_1}^{f_1} \sum \left(C_{\hat{f}_2}^{f_2} \dots C_{\hat{f}_m}^{f_m} \right)}{C_{N+m-1}^{N-n}} = \frac{C_{\hat{f}_1}^{f_1} C_{N+m-2-\hat{f}_1}^{N-n-\hat{f}_1+f_1}}{C_{N+m-1}^{N-n}}$$

$$\frac{C_{\hat{f}_i}^{f_i} C_{N+m-2-\hat{f}_i}^{N-n-\hat{f}_i+f_i}}{C_{N+m-1}^{N-n}} \xrightarrow{N\to\infty} \frac{(n+m-1)!}{f_i!\,(n-f_i+m-2)!}\pi_i^{f_i}(1-\pi_i)^{n-f_i+m-2}$$

$$\frac{(n+m-1)!}{f_i!\,(n-f_i+m-2)!}\pi_i^{f_i}(1-\pi_i)^{n-f_i+m-2} \xrightarrow{n\to\infty} N(\mu;\sigma^2)$$

$$\mu = \frac{f_i}{n+m-2} \qquad \xrightarrow{m=2} \qquad \mu = \frac{f_i}{n} = p$$

$$\sigma^2 = \frac{\mu(1-\mu)}{n+m-2} \qquad \qquad \sigma^2 = \frac{p(1-p)}{n}$$

k_haralampiev@phls.uni-sofia.bg

http://kaloyan-haralampiev.info