A MODIFICATION OF BAYES' PROBLEM

By R. v. Mises

The classical Bayes problem can be stated as follows. We consider an urn which contains white and black balls (or balls designated by 0 and 1). The probability p for drawing a black ball is unknown. But there is given a probability function $F(x)$ representing the *a priori* probability for the inequality $p \leqq x$. We draw n times from the urn (returning each time the extracted ball) and get a black ball m times and a white one $n - m$ times. Now, after this experiment, we ask for the *a posteriori* probability $P_n(x)$ for the relation $p \leqq x$.

The solution proposed by Bayes can be written in a slightly generalized form:

$$P_n(x) = K \int_0^x p^m (1-p)^{n-m} \, dF(p) \tag{1}$$

where K is a constant to be found by means of the condition

$$P_n(1) = 1. \tag{1'}$$

We are interested in the behaviour of $P_n(x)$ if n tends to ∞ under the condition

$$\lim_{n\to\infty} \frac{m}{n} = \alpha. \tag{2}$$

Laplace found in the case of a priori equipartition $F(x) = x$, and I proved in 1919[1] for any derivable $F(x)$, that $P_n(x)$ tends to a normal distribution:

$$\lim_{n\to\infty} \left[P_n(x) - \frac{1}{\sqrt{\pi}} \int_{-\infty}^{u} e^{-u^2} \, du \right] = 0 \tag{3}$$

with

$$u = H_n(x - A_n)$$

$$A_n = \alpha, \qquad \frac{1}{2H_n^2} = \frac{\alpha(1-\alpha)}{n}. \tag{4}$$

It is easily seen from (3) and (4) that

$$\lim_{n\to\infty} P_n(x) = \begin{cases} 0 & \text{if } x < \alpha \\ 1 & \text{`` } x > \alpha. \end{cases} \tag{5}$$

Let us now consider a slightly modified form of the problem.[2] Instead of one

---

[1] *Mathematische Zeitschrift*, vol. 4 (1919) p. 92. Cf. my textbook *Wahrscheinlichkeitsrechnung und ihre Anwendungen*, Wien-Leipzig 1931, p. 158. Later I proved the Laplace-Bayes theorem for a more general class of $F(x)$: *Monatshefte für Mathematik und Physik*, vol. 43 (1936) pp. 105–128.

[2] This modified problem has been treated by S. Bochner, *Annals of Math.*, Vol. 37, 1936, p. 816.

urn we suppose there are given n urns each containing white and black balls. The probability p_ν for drawing a black ball from the ν^{th} urn is unknown, but is subject to an a priori probability function $F(x)$ which furnishes the a priori probability for the relation $p_\nu \leqq x$, independently of ν. We assume that on drawing one ball from every urn a black ball appears m times and a white ball $n - m$ times. Putting

$$\frac{p_1 + p_2 + \cdots + p_n}{n} = p, \tag{6}$$

we ask for the a posteriori probability $P_n(x)$ for the relation $p \leqq x$.

The Bayes formula (1) must now be replaced by

$$\begin{aligned} P_n(x) = K' \int\int_{p_1+p_2+\cdots p_n \leqq nx} \cdots \int p_1 p_2 \cdots p_m (1 - p_{m+1})(1 - p_{m+2}) \\ \cdots (1 - p_n)\, dF(p_1) \cdots dF(p_n) \end{aligned} \tag{7}$$

where K' is a constant determined by (1'). It is very easy to examine the asymptotic character of (7). We shall prove the following

THEOREM: *If the first three moments of the a priori distribution $F(x)$*

$$b_\nu = \int_0^1 x^\nu\, dF(x), \qquad \nu = 1, 2, 3 \tag{8}$$

exist and if the dispersion $b_2 - b_1^2$ is different from 0, the a posteriori probability $P_n(x)$ tends for $n \to \infty$ under the condition (2) to the normal distribution (3) with

$$\begin{aligned} A_n &= \alpha \frac{b_2}{b_1} + (1 - \alpha) \frac{b_1 - b_2}{1 - b_1} \\ \frac{1}{2H_n^2} &= \frac{1}{n}\left[\alpha \frac{b_1 b_3 - b_2^2}{b_1^2} + (1 - \alpha) \frac{(b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2}{(1 - b_1)^2}\right]. \end{aligned} \tag{9}$$

In order to prove the theorem we write

$$\begin{aligned} V_\nu(p_\nu) &= \frac{1}{b_1}\int_0^{p_\nu} x\, dF(x), \quad \text{if } \nu = 1, 2, \cdots m \\ &= \frac{1}{1 - b_1}\int_0^{p_\nu} (1 - x)\, dF(x), \quad \text{if } \nu = m + 1, m + 2, \cdots n. \end{aligned} \tag{10}$$

Then formula (7) becomes

$$P_n(x) = C \int\int_{p_1+p_2+\cdots p_n \leqq nx} \cdots \int dV_1(p_1)\, dV_2(p_2) \cdots dV_n(p_n). \tag{11}$$

Each $V_\nu(p_\nu)$ is a distribution function, i.e. a non-decreasing function with $V_\nu(-\infty) = 0$, $V_\nu(\infty) = 1$. Therefore the constant C in (11) is equal to 1 and the integral represents the distribution function for the arithmetical mean $(p_1 + p_2 + \cdots p_n)/n$. According to the *Central Limit Theorem* of the theory of probability $P_n(x)$ will converge towards a normal distribution when certain conditions are satisfied. In every case, if a_ν, s_ν^2 denote the mean value and the dispersion associated with $V_\nu(x)$, then the mean value A_n and the dispersion S_n^2 associated with $P_n(x)$ will be defined by

$$A_n = \frac{1}{n}\sum_{\nu=1}^{n} a_\nu, \qquad S_n^2 = \frac{1}{n^2}\sum_{\nu=1}^{n} s_\nu^2. \tag{12}$$

We find from (10)

$$\begin{aligned} a_\nu &= \int_0^1 x\,dV_\nu(x) = \frac{1}{b_1}\int_0^1 x^2\,dF(x) = \frac{b_2}{b_1}, \quad \text{if } \nu = 1, 2, \cdots m \\ &= \frac{1}{1-b_1}\int_0^1 x(1-x)\,dF = \frac{b_1 - b_2}{1-b_1}, \quad \text{if } \nu = m+1, \cdots n \end{aligned} \tag{13}$$

$$\begin{aligned} s_\nu^2 &= \int_0^1 x^2\,dV_\nu(x) - a_\nu^2 = \frac{b_3}{b_1} - \frac{b_2^2}{b_1^2}, \quad \text{if } \nu = 1, 2, \cdots m \\ &= \frac{b_2 - b_3}{1-b_1} - \frac{(b_1 - b_2)^2}{(1-b_1)^2}, \quad \text{if } \nu = m+1, \cdots n. \end{aligned} \tag{14}$$

We supposed the dispersion of $F(x)$ to be different from zero. It follows that

$$b_1 \neq 0,\ 1 - b_1 \neq 0,\ b_3b_1 - b_2^2 \neq 0,\ (b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2 \neq 0. \tag{15}$$

For $b_1 = 0$ would imply that $dF(x) = 0$ for all $x > 0$ and $1 - b_1 = 0$ that $dF(x) = 0$ for all $x < 1$; in both cases the dispersion would be zero. On the other hand, it is easily seen that the relation $b_3b_1 - b_2^2 = 0$ is not compatible with the condition of a non-vanishing a priori dispersion and that the same is true for the relation $(b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2 = 0$.

The total dispersion Σs_ν^2 is equal to the sum of m times the value $(b_3b_1 - b_2^2)/b_1^2$ and $n - m$ times the value $[(b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2]/(1 - b_1)^2$.

Thus we see that under the condition (2) the sum Σs_ν^2 tends to ∞, while the ratio $s_\nu^2/\Sigma s_\nu^2$ tends to zero, if n increases infinitely. These are sufficient conditions for the validity of the Central Limit Theorem.[3] The values given for A_n and H_n^2 in (9) follow from (12), (13), (14) and the well known relation $2H_n^2S_n^2 = 1$.

S. Bochner in his previously quoted paper found, in a more complicated manner, the value of A_n and only showed that $P_n(x)$ tends to zero if $x < A_n$ and to 1 if $x > A_n$.

EXAMPLES. If we assume the a priori probability to be uniform, i.e. $F(x) = x$, we have

$$b_1 = \tfrac{1}{2}, \qquad b_2 = \tfrac{1}{3}, \qquad b_3 = \tfrac{1}{4}$$

and therefore from (9)

[3] Cf. H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tract in Mathematics and Mathematical Physics, No. 36, 1937, p. 56.

$$A_n = \tfrac{1}{3}(\alpha + 1), \qquad \frac{1}{2H_n^2} = \frac{1}{18n}.$$

A more general case is that of a more concentrated a priori probability function

$$F'(x) = Cx^k(1-x)^l, \qquad C = \frac{(k+l+1)!}{k!\,l!}.$$

Here we find

$$b_1 = \frac{k+1}{k+l+2}, \qquad b_2 = \frac{(k+1)(k+2)}{(k+l+2)(k+l+3)},$$

$$b_3 = \frac{(k+1)(k+2)(k+3)}{(k+l+2)(k+l+3)(k+l+4)}$$

and the values of A_n and H_n^2 are

$$A_n = \frac{\alpha+k+1}{k+l+3}, \qquad \frac{1}{2H_n^2} = \frac{\alpha(l-k)+(k+1)(l+2)}{n(k+l+3)^2(k+l+4)}.$$

By introducing the moments of $F(x)$ relative to the mean value, i.e.

$$(16) \qquad \begin{aligned} B_2 &= \int_0^1 (x-b_1)^2\,dF = b_2 - b_1^2, \\ B_3 &= \int_0^1 (x-b_1)^3\,dF = b_3 - 3b_1b_2 + 2b_1^3 \end{aligned}$$

we can transform the general formulas (9) into

$$(17) \qquad \begin{aligned} A_n &= b_1 + \frac{B_2}{b_1(1-b_1)}(\alpha - b_1) \\ \frac{1}{2H_n^2} &= \frac{1}{n}\left[B_2 + B_3\frac{\alpha-b_1}{b_1(1-b_1)} - B_2^2\frac{b_1^2+\alpha(1-2b_1)}{b_1^2(1-b_1)^2}\right]. \end{aligned}$$

The first of these equations shows that the a posteriori mean value A_n (for all n) is equal to the a priori mean value b_1, if the experimental mean m/n or α coincides with the latter. On the other hand, in the case of a symmetric a priori distribution ($b_1 = \frac{1}{2}$, $B_3 = 0$) the second equation is reduced to

$$\frac{1}{2H_n^2} = \frac{1}{n}(B_2 - 4B_2^2).$$

On the whole it is remarkable that the influence of the a priori probability does not vanish for $n \to \infty$, in the case of our modified Bayes problem.[4] The explanation of this fact is to be found in a more generalized theory of the inverse problems in probability.

UNIVERSITY OF ISTANBUL, TURKEY.

[4] Cf. my papers quoted in footnote 1.