# Estimation of a Bernoulli Parameter p from Imperfect Trials

Petar M. Djurić and Yufei Huang

Abstract—Imperfect Bernoulli trials arise when the outcome of a Bernoulli experiment is not known with certainty. In signal processing, we often need to estimate a probability of occurrence p of an event from imperfect Bernoulli trials. A typical example is the estimation of the probability of a signal being present in noisy data. In his famous essay, Bayes solved the same problem but for perfect trials. In this letter, a solution is provided for imperfect trials. It is shown that it includes Bayes' solution as a special case.

*Index Terms*—A posteriori densities, estimation, probabilities of events, signal detection.

# I. INTRODUCTION

N his famous work, "An essay toward solving a problem in the doctrine of chance," Thomas Bayes addressed the following problem: "Given: the number of times in which an unknown event had happened and failed: Required: the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named" [1]. In mathematical terms, if p is the probability that a phenomenon of interest may occur, m is the number of times a Bernoulli experiment is conducted, and  $k, k \le m$  is the number of times the event occurred, Bayes wanted to find the probability  $P(a , where <math>0 \le a < b \le 1$ , and  $\mathcal{D}$  generically denotes experimental data. Bayes solved the problem without relying on integral calculus and instead used interesting geometric arguments. It should be noted that he tacitly assumed that the outcomes of the Bernoulli experiments are known with certainty. That is, after observing the data from the trial, the probability of occurrence of the event would collapse to either zero or one. Such Bernoulli trials are here referred to as perfect Bernoulli trials.

In signal processing, there are many situations where Bernoulli experiments are performed and the value of the probability of occurrence of an event is desired. A typical scenario there, however, is different from the one described by Bayes in that the data do not reveal with certainty whether the event occurred. For example, consider a problem where there are mindependent data records and each of them represents a signal in noise or noise only, and one wants to estimate from past experiments the probability that a signal will be present in the next data set. For each data record, two hypotheses are formed

The authors are with the Department of Electrical and Computer Engineering, State University of New York, Stony Brook, NY 11794-2350 USA (e-mail: djuric@ece.sunysb.edu).

Publisher Item Identifier S 1070-9908(00)05105-1.

and their posterior probabilities are evaluated. Theoretically, due to the presence of noise, the posterior probabilities of each of the hypotheses are never zero or one (i.e., for each data record, the two posterior probabilities are always nonzero, and so the application of Bayes' solution cannot be directly applied). This is particularly critical when the signal is weak and the posterior probabilities of the hypotheses take values close to 1/2. To distinguish this setting from the one solved by Bayes, we refer to Bernoulli experiments whose outcomes are not known with certainty as imperfect Bernoulli trials. For some reason, this problem has not received much attention in the literature. The only direct reference we found on the subject is [4], where a scenario that arises in social sciences is examined. The objective of this letter is to present the Bayes' solution for imperfect Bernoulli trials and show that the solution for perfect trials is a special case of it.

### II. ESTIMATION FROM PERFECT BERNOULLI TRIALS

Here, we briefly review the case of perfect Bernoulli trials [3]. Without loss of generality, in the sequel, we assume that a = 0 and b = 1. Let p be the probability that an event Q will occur in a Bernoulli trial, that is P(Q) = p, where p is the parameter we want to estimate. Suppose that the Bernoulli experiment is repeated m times under identical conditions, and that the observed data from the trials are denoted by  $\mathbf{y}_i$ ,  $i = 1, 2, \dots, m$ . The data  $\mathbf{y}_i$  belong to one of two disjoint sets, S and  $\overline{S}$ , and we say that Q occurred in the *i*th trial if  $\mathbf{y}_i \in S$ . Therefore

$$P(\mathbf{y}_i \in S) = p$$
  

$$P(\mathbf{y}_i \in \overline{S}) = 1 - p.$$
(1)

It is important to emphasize that after observing  $\mathbf{y}_i$ , we can unambiguously decide if  $\mathbf{y}_i \in S$  or  $\mathbf{y}_i \in \overline{S}$ . If in *m* trials *Q* occurs *k* times, then for the probability of  $\mathbf{y}_i$ ,  $i = 1, 2, \dots, m$ , we can write

$$P(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m | p) = p^k (1-p)^{m-k}.$$
 (2)

If f(p) is our prior of p, according to the Bayes theorem, the posterior density of p is

$$f(p|\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m) \propto P(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m|p)f(p) \quad (3)$$

where  $\propto$  signifies proportionality. From (2), the posterior (3) becomes

$$f(p|\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m) \propto p^k (1-p)^{m-k} f(p).$$
(4)

1070-9908/00\$10.00 © 2000 IEEE

Manuscript received September 4, 1999. This work was supported by the National Science Foundation Award CCR-9903120. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. A. Nehorai.

When the prior is uniform on [0, 1], the posterior becomes [3]

$$f(p|\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m) = \frac{(m+1)!}{k!(m-k)!} p^k (1-p)^{m-k}$$
(5)

which is the well known *beta density* [2]. This result can be found in many standard textbooks on probability and statistics.

#### **III. ESTIMATION FROM IMPERFECT BERNOULLI TRAILS**

Now let each Bernoulli experiment produce a data record  $\mathbf{y}_i$ ,  $i = 1, 2, \dots, m$  that can be modeled by one of two hypotheses,  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , and let the densities of the data given these hypotheses be  $f(\mathbf{y}_i|\mathcal{H}_0)$  and  $f(\mathbf{y}_i|\mathcal{H}_1)$ , respectively. If the prior of p is f(p), after observing the first data record  $\mathbf{y}_1$ , we can write

$$f(p|\mathbf{y}_{1}) \propto f(\mathbf{y}_{1}|p)f(p)$$

$$= f(\mathbf{y}_{1}|p,\mathcal{H}_{0})P(\mathcal{H}_{0}|p)f(p)$$

$$+ f(\mathbf{y}_{1}|p,\mathcal{H}_{1})P(\mathcal{H}_{1}|p)f(p)$$

$$= f(\mathbf{y}_{1}|\mathcal{H}_{0})P(\mathcal{H}_{0}|p)f(p)$$

$$+ f(\mathbf{y}_{1}|\mathcal{H}_{1})P(\mathcal{H}_{1}|p)f(p).$$
(6)

The last identity follows because if the density of  $y_1$  is conditioned on p and  $\mathcal{H}_i$ , the conditioning on p becomes irrelevant. Since

$$p = P(\mathcal{H}_1|p) \tag{7}$$

and

$$1 - p = P(\mathcal{H}_0|p) \tag{8}$$

(6) can be expressed as

$$f(p|\mathbf{y}_1) \propto ((1-r_1)p + r_1)f(p)$$
 (9)

where

$$r_1 = \frac{f(\mathbf{y}_1 | \mathcal{H}_0)}{f(\mathbf{y}_1 | \mathcal{H}_1)}.$$
(10)

The prior for the second data record is given by (9), and the posterior is

$$f(p|\mathbf{y}_1, \mathbf{y}_2) \propto ((1-r_2)p + r_2)f(p|\mathbf{y}_1)$$
 (11)

where

$$r_2 = \frac{f(\mathbf{y}_2|\mathcal{H}_0)}{f(\mathbf{y}_2|\mathcal{H}_1)} \tag{12}$$

or

$$f(p|\mathbf{y}_1, \mathbf{y}_2) \propto f(p) \prod_{i=1}^2 ((1-r_i)p + r_i).$$
 (13)

It is easy to show that the result for the posterior after m trials is

$$f(p|\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m) \propto f(p) \prod_{i=1}^m \left( (1-r_i)p + r_i \right)$$
(14)

where

$$r_i = \frac{f(\mathbf{y}_i|\mathcal{H}_0)}{f(\mathbf{y}_i|\mathcal{H}_1)}.$$
(15)

When f(p) is uniform on [0, 1], and provided none of the trials produces  $r_i = 1$ , the posterior can be written as

$$f(p|\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m) = a_{m, m} p^m + a_{m, m-1} p^{m-1} + \dots + a_{m, 0}$$
(16)

where the coefficients of the polynomial  $a_{m,i}$ ,  $i = 0, 1, \dots, m$ are defined by

$$a_{m,i} = \frac{c_{m,i}}{\sum_{k=0}^{m} \frac{c_{m,k}}{k+1}}$$
(17)

and where  $c_{m,i}$ ,  $i = 0, 1, \dots, m$  can be obtained from

$$c_{m,0} = c_{m-1,0} r_m \tag{18}$$

$$c_{m,i} = c_{m-1,i-1}(1 - r_m) + c_{m-1,i}r_m,$$

$$i = 1, 2, \cdots, m-1$$
 (19)

$$c_{m,m} = c_{m-1,m-1}(1 - r_m) \tag{20}$$

and

$$c_{1,0} = r_1$$
 (21)

$$c_{1,1} = 1 - r_1. \tag{22}$$

So if the exact analytical form of  $f(p|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)$  is desired, the coefficients  $a_{m,i}$  that define the density can recursively be obtained from (17)–(22).

From the above, we see that  $f(p|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)$  is a polynomial of *m*th degree provided none of the ratios  $r_i$  equal one. If n of the ratios are equal to one, then the polynomial is of degree m - n. In perfect Bernoulli trials,  $r_i$  is either 0 or  $\infty$ . When it is zero, the factor in (14) from the *i*th trial is p, whereas if it is  $\infty$ , the factor is (1-p). These factors, when multiplied and the product normalized, yield Bayes' solution (5).

To get further insight about the posterior obtained from imperfect trials, let  $P(\mathcal{H}_1|\mathbf{y}_i) = p^{(i)}$ , i = 1, 2, denote the posteriors that hypothesis  $\mathcal{H}_1$  is true after observing  $\mathbf{y}_i$ . Then we may write

$$f(p|\mathbf{y}_1) = (4p^{(1)} - 2)p + 2 - 2p^{(1)}$$
(23)

and

$$f(p|\mathbf{y}_1, \mathbf{y}_2) = a_{2,2}p^2 + a_{2,1}p + a_{2,0}$$
(24)

where

$$a_{2,2} = \frac{6(2p^{(1)} - 1)(2p^{(2)} - 1)}{2p^{(1)}p^{(2)} - p^{(1)} - p^{(2)} + 2}$$
(25)  
$$a_{2,1} = \frac{6((2p^{(1)} - 1)(1 - p^{(2)}) + (2p^{(2)} - 1)(1 - p^{(1)}))}{2p^{(1)}p^{(2)} - 1p^{(1)}p^{(1)}}$$

$$\frac{2p^{(1)}p^{(2)} - p^{(1)} - p^{(2)} + 2}{(26)}$$

$$a_{2,0} = \frac{6(1-p^{(1)})(1-p^{(2)})}{2p^{(1)}p^{(2)}-p^{(1)}-p^{(2)}+2}.$$
(27)

The effects of  $p^{(1)}$  and  $p^{(2)}$  on the posterior of p are shown in Figs. 1 and 2. In Fig. 1, we see that if  $p^{(1)} = 1$  or  $p^{(1)} = 0$ , we have the beta density and the Bayes' solution. If  $p^{(1)} = 0.5$ , we notice that the posterior has not changed. That is, it is still flat because then the evidence from data record  $y_1$  favors neither

Authorized licensed use limited to: SUNY AT STONY BROOK. Downloaded on April 30,2010 at 17:35:44 UTC from IEEE Xplore. Restrictions apply.



Fig. 1. Posterior probability of p after one data record for various probabilities  $p^{(1)}$  (probability that  $\mathcal{H}_1$  is true given the data record  $\mathbf{y}_1$ ).



Fig. 2. Posterior probability of p after two data records for various probabilities  $p^{(2)}$  (probability that  $\mathcal{H}_1$  is true given the data record  $\mathbf{y}_2$ ) and  $p^{(1)} = 0.8$  (probability that  $\mathcal{H}_1$  is true given the data record  $\mathbf{y}_1$ ).

 $\mathcal{H}_1$  nor  $\mathcal{H}_0$ . For all other values of  $p^{(1)}$ , the function  $f(p|\mathbf{y}_1)$  is accordingly slanted. In Fig. 2,  $f(p|\mathbf{y}_1, \mathbf{y}_2)$  is plotted for  $p^{(1)} = 0.8$  and for various values of  $p^{(2)} \in [0, 1]$ . As  $p^{(2)}$  varies from 0 to 1, the shape of the posterior changes considerably. For  $p^{(2)} = 0.5$ , we see that the posterior is a linear function of p because, as has already been noted, when the ratio  $r_i = 1$  (which is equivalent to  $p^{(i)} = 0.5$ ), new evidence of p is nonexistent, and the function of the posterior does not increase the degree of its polynomial.

# IV. AN EXAMPLE

In this section, examples are provided that show the difference between Bayesian estimates from imperfect Bernoulli trials when they are obtained by considering the trials perfect and imperfect, respectively. Namely, one attempt in estimating p by treating the trials as "perfect" could be based on applying (5), that is, increasing the count k by one whenever



Fig. 3. Posterior density of p obtained with the method for imperfect Bernoulli trials (solid line) and the method for perfect Bernoulli trials (dotted line), where p = 0.7. The number of trials m = 200.

 $r_i < \gamma$ , where  $\gamma$  is an appropriately chosen threshold (in our experiment, we adopted  $\gamma = 1$ ). Then, it is well known that for large m, the function  $\phi(p) = p^k(1-p)^{m-k}$  has a sharp peak at k/m, and so, if the prior p is smooth, the posterior  $f(p|\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m)$  will also be concentrated around k/m [3].

In our experiment, each time a trial is performed, a data vector  $y_i$  of length T = 30 samples is observed. It is generated by one of the hypotheses

$$\mathcal{H}_0: y[t] = w[t], \qquad t = 0, 1, \cdots, T - 1$$
  
$$\mathcal{H}_1: y[t] = a + w[t], \qquad t = 0, 1, \cdots, T - 1 \qquad (28)$$

where the w[t]'s are independent and identically distributed noise samples whose distribution is Gaussian with mean zero and known variance  $\sigma^2$ . The signal *a* is random and also Gaussian with mean zero and known variance  $\sigma_a^2$ .

The experiment had m = 200 trials, with  $p(\mathcal{H}_1) = 0.7$ . Each trial was conducted as follows: First a random number u was drawn from a uniform distribution,  $\mathcal{U}(0, 1)$ . Second, if the number was less than 0.7, y was generated according to  $\mathcal{H}_1$ ; otherwise it was generated according to  $\mathcal{H}_0$ . If  $\mathcal{H}_1$  was the generating mechanism, a was first sampled from  $\mathcal{N}(0, \sigma_a^2)$  followed by generating  $\mathbf{w} \sim \mathcal{N}(0, \mathbf{I})$ . The variance of a was  $\sigma_a^2 = 1/30$ .

The ratio  $r_i$ , for  $i = 1, 2, \dots, m$ , can be shown to be [2]

$$r_i = \left(\frac{\sigma^2}{\sigma^2 + T\sigma_a^2}\right)^{1/2} \exp\left(\frac{T^2 \overline{y}_i^2 \sigma_a^2}{2(T\sigma_a^2 + \sigma^2)\sigma^2}\right)$$
(29)

where  $\overline{y}_i$  is the mean of  $\mathbf{y}_i$ . With the method for perfect Bernoulli trials, whenever  $r_i < 1$ , it was considered that the data were produced according to hypothesis  $\mathcal{H}_1$ , and if  $r_i \geq 1$ , it was supposed that they were generated by  $\mathcal{H}_0$ .

The results of the experiment are shown in Fig. 3. The method for perfect Bernoulli trials obviously failed to provide accurate result because it does not incorporate the uncertainties present in the data when deciding between  $\mathcal{H}_1$  or  $\mathcal{H}_0$ . The posterior obtained by applying (14) is shown by the curve with solid line,



Fig. 4. Posterior density of p obtained with the methods for imperfect Bernoulli trials (solid line) and the method for perfect Bernoulli trials (dotted line), where p = 0.7. The number of trials m = 30, and the SNR's were -20 dB (top), -10 dB (middle), and 0 dB (bottom figure).

and it peaks at about the correct value of p. In general, the posterior densities of p obtained according to the imperfect Bernoulli formulation are always wider than those obtained by the method based on perfect Bernoulli trials.

A different set of experiments were performed, where a was kept constant and equal to one, and the noise variance was varied from experiment to experiment. The total number of records

in each experiment was m = 30. In Fig. 4, we see the results for SNR's (SNR's) -20, -10, and 0 dB, respectively ( $SNR = 10 \log a^2/\sigma^2$ ). When the SNR was 0 dB, the found posteriors by the methods for perfect and imperfect Bernoulli trials were practically the same (there is no distinction between the two curves in the bottom graph of Fig. 4), and as the SNR decreased, the posterior obtained by the method for imperfect trials became broader. This certainly agrees with our intuition and simply reflects the fact that the noisier data induce greater uncertainty about the estimated probability.

# V. CONCLUSIONS

Estimation of the probability of an event from imperfect Bernoulli trials was examined. The main result is given by (14)–(22), which represents the posterior of the probability given data from m independent and imperfect Bernoulli trials. The expression simplifies to (5) when the trials are perfect. From the posterior of the probability, various point and interval estimates of the probability of the event can be readily obtained.

#### REFERENCES

- T. Bayes, "An essay toward solving a problem in the doctrine of chances," *Philos. Trans. Roy. Soc. Lond.*, vol. 53, pp. 370–418, 1763.
- [2] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York: Wiley, 1994.
- [3] A. Papoulis, Probability, Random Variables, and Stochastic Processes. New York: McGraw-Hill, 1991.
- [4] R. L. Winkler and A. Gaba, "Inference with imperfect sampling from a Bernoulli process," in *Bayesian and Likelihood Methods in Statistics and Econometrics*, S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner, Eds. Amsterdam, The Netherlands: North Holland, 1990.