TEACHING OF BAYESIAN ESTIMATION OF "P" PROBABILITY IN A BERNOULLI PROCESS

María Teresa Gil
Universidad Nacional Buenos Aires, Argentina
tgil@fi.uba.ar

Both at the School of Engineering of Universidad Nacional de La Matanza and Universidad de Buenos Aires, the teaching of statistics is considered to be highly important. The aim of this presentation is to show the advantages and disadvantages of the Bayesian estimation of the "p" parameter in a Bernoulli Process, in particular in a binomial distribution (n ; p). The similarities and differences between the classical estimation of such parameter and the Bayesian estimation will be established. The predictions which can be carried out as a result of the obtained estimations will also be developed. In order to teach both Bayesian and classical estimation of population parameters, mathematical and statistics software is used as a tool to help the understanding of the concepts. In this presentation Matlab *and* Excel *have been used.*

INTRODUCTION

In the teaching of statistics it is important that the learners clearly understand both the differences and similarities between the conventional and Bayesian paradigms for different reasons:

- unlike the classical estimation method, the Bayesian estimation method incorporates initial information;
- it will help the learner to decide which method is the most convenient to be applied in specific problems regarding Engineering or any other subject.

Following a short description of Bayesian estimation of "*p*" in a Binomial (*n*; *p*), examples of how to estimate in a Bayesian way the probability of defective integrated circuits (I.C.) are presented. The results of these estimations are then compared to the results obtained through the classical estimation method.

BAYESIAN ESTIMATION OF "*p*" IN A BINOMIAL (*n* ; *p*)

- A function of prior uniform density of "*p*" is considered [0 ; 1], i.e., Beta (1 ; 1)
- A single sample is taken
- "*r*" successes in "*n*" Bernoulli experiments are observed (dichotomic)
- Therefore, the posterior density function of "*p*" is:

$$BETA \quad (\alpha = r+1 \quad ; \quad \beta = n-r+1)$$

$$\mu = \frac{r+1}{n+2} \qquad mo_p = \frac{r}{n} = \hat{p}_{MV}$$

It is observed the following *similarity* between the classical estimation of such parameter and the Bayesian estimation:

the mode of the posterior density function of "p" in this case coincides with the maximum likelihood estimator of "p."

CONFIDENTIAL INTERVAL TO ESTIMATE A "*p*"

A credibility interval of (1-α) % credibility to estimate (in Bayesian way) "*p*" is sought, i.e.,

$$P[LI \le p \le LS] = 1-\alpha$$

The limits of such interval can be obtained, for instance stating that:

$$\int_0^{LI} f\left(p/_{r,n}\right) dp = \frac{\alpha}{2} \qquad \int_0^{LS} f\left(p/_{r,n}\right) dp = 1-\frac{\alpha}{2}$$

Example 1: In the following example two cases are shown in order to estimate, in a Bayesian way, the "*p*" probability of defective integrated circuits (I. C.), and are compared to the results obtained through the conventional estimation method. In these two cases, there is a *similarity* between a credibility interval for "p" obtained in a classical way and in a Bayesian way when a *big sample* is taken.

In order to achieve this, in each of the two cases:

- A uniform distribution [0; 1] as prior distribution of "*p*" is assumed.
- A single sample of "*n*" I.C. in which "*r*" defective ones are observed is taken.
- The intervals of 95% credibility in each of the two cases are:

Bayesian Estimation		*Conventional Estimation*	
$n = 100$	$r = 10$	$n = 100$	$r = 10$
Mean of posterior density function = 0.1078			
Mode of posterior density function = 0.1		Maximum likelihood estimation of "*p*" = 0,1	
$L_I = 0.0556$	$L_S = 0.1746$	$L_I = 0.0412$	$L_S = 0.1588$
$n = 1000$	$r = 10$	$n = 1000$	$r = 10$
Mean of posterior density function = 0.0110			
Mode of posterior density function = 0.01		Maximum likelihood estimation of "*p*" = 0.01	
$L_I = 0.0055$	$L_S = 0.0183$	$L_I = 0.0038$	$L_S = 0.01617$

Example 2: In this example the *difference* between a credibility interval for "*p*" obtained in a classical way and in a Bayesian way is shown.

- A uniform distribution [0;1] as prior "*p*" distribution is assumed.
- A single sample of 100 I.C. in which one defective is observed is taken.
- The intervals of 95% credibility are:
  LI Bayes = 0.0024 LS Bayes = 0.0539
  LI conventional = *-0.0095* LS conventional = 0.0295

It is observed that the interval obtained in a classical way produces as inferior limit a negative value, which is impossible as likelihood value of defective I.C.

Example 3: The following example shows two cases to estimate (in a Bayesian way) the "*p*" likelihood of defective I.C.
In order to achieve this in each of the cases:

- A Uniform distribution [0;1] as prior distribution of "*p*" is assumed.
- A single sample of 100 I.C. is taken, in which in the first case "*r*=0" defective ones and in the second "*r*=100" defective ones are observed.
- The interval of 95% credibility in the first case "*r*=0" is: LI = 5.0773 e-4 LS = 0.0359

In this case it is more logical to provide a unilateral credibility interval of 95% to estimate "*p*," i.e.,: LI = 0 LS = 0.0292

- The interval of 95% credibility in the second case "*r*=100" is: LI = 0.9708 LS = 0.9997

In this case it is more logical to provide a unilateral credibility interval of 95% to estimate "*p*," i.e.,: LI = 0.9995 LS = 1

It is worth pointing out that in these cases a credibility interval for "p" cannot be found by means of the classical estimation.

PREDICTIVE FUNCTION OF THE BINOMIAL (n ; p)

The aim is to search the likelihood function of the "*r*" amount of defective I. C. conditional on the first obtained sample in order to predict the likelihood of obtaining "*r*" defective I.C. in a second sample. The predictive function is: Beta-Binomial:

$$Bb\left(r \middle/ \alpha = r_1 + 1; \beta = n_1 - r_1 + 1; n_2 \right) \qquad P\left(r \middle/ n_2 ; n_1 ; r_1 \right) = \frac{\binom{n_2}{r}\binom{n_1}{r_1}}{\binom{n_2 + n_1}{r + r_2}} \quad for\ r = 0,1,2,3,....n_2$$

Examples: A first sample has been taken of n_1=50 I.C. obtaining r_1=1 defective ones. The posterior function of "*p*" has been calculated (which provides information about all the possible values of "*p*") The aim is to predict the likelihood of obtaining "*r*" defective I.C. in a posterior sample of n_2=10 I.C.
Therefore:

	Bayesian estimation	*Conventional estimation*
r	*P(r / n2,n1,r1)*	*P(r / n2; p_{MLE})*
0	0.696721311	0.817072807
1	0.236176716	0.166749552
2	0.054972167	0.015313734
3	0.010287189	0.000833401
4	0.001607373	2.97643E-05
5	0.00021042	7.28922E-07
6	2.27305E-05	1.23966E-08
7	1.96058E-06	1.44567E-10
8	1.2725E-07	1.10638E-12
9	5.54464E-09	5.0176E-15
10	1.21982E-10	1.024E-17

A second sample has been taken of $n_1 = 50$ $r_1 = 0$ Second sample: $n_2 = 10$

	Bayesian estimation	*Conventional estimation*
r	*P(r / n2,n1,r1)*	*P(r / n2, p_{MLE})*
0	0.836065574	1
1	0.139344262	0
2	0.021255904	0
3	0.002931849	0
4	0.000360052	0
5	3.8577E-05	0
6	3.507E-06	0
7	2.59778E-07	0
8	1.47044E-08	0
9	5.65553E-10	0
10	1.10893E-11	0

In this second example it is observed that the *advantage* of the Bayesian estimation of the "*p*" parameter in a binomial distribution (*n* ; *p*)

POSTERIOR DENSITY FUNCTION OF "p" WITH "M" SAMPLES OF A BINOMIAL (*n* ; *p*) AND PRIOR BETA (*a* ; *b*)

It can be proved in this process that if M samples $\{r_i\}_{i=0}^{M-1}$ are taken from a Binomial (n ; p), the posterior distribution of "*p*" is the same as if only one sample of size "*M*×*n*" is taken and are observed $r_M = \sum_{i=0}^{M-1} r_i$ (the sample mean will be $\overline{r} = r_M / M$).

$$f\left(p / \underline{r}_M\right) = \frac{P\left(\underline{r}_M / p\right) f(p)}{\int_0^1 P\left(\underline{r}_M / p\right) f(p)\, dp}$$

If a prior beta likelihood distribution is considered, $Beta \ \ (\alpha = a, \ \beta = b)$, the function of posterior likelihood density is: $Beta \ \ (\alpha = M\overline{r} + a \ ; \ \beta = M\left(n - \overline{r}\right) + b)$

$$mean = E\left(p / \underline{r}_M\right) = \frac{M\overline{r} + a}{Mn + a + b} \qquad\qquad mo_p = \frac{M\overline{r} + a - 1}{Mn + a + b - 2}$$

Remark 1: If a Beta distribution as prior of "*p*" is considered, the posterior distribution is also Beta. This enables to make an iterative estimation, taking the posterior Beta distribution of "*p*" of iteration "*i*," as the prior distribution of "*p*" in iteration "*i*+1." This is very useful in cases where it is difficult to take big samples in one go and it also allows to process *on line* the samples as they arrive.

Remark 2: The *similarity* between the classical estimation of such parameter and the Bayesian estimation, is: when $M \rightarrow \infty$, $E\left(p / \underline{r}_M\right) \rightarrow \frac{\overline{r}}{n} = p_{mv}$.

CONCLUSION

With simple examples like the ones presented in this paper, one can show the *similarities* and *differences* between the classical estimation of population parameters and the Bayesian estimation, as well as the *predictions* which can be carried out as a result of the obtained estimations. It is important to teach both estimation methods and how they compare, so that the learner can apply them in real situation.

REFERENCES

Bernardo, J. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley.

Canavos, G. C. (1988). *Probabilidad y Estadística Aplicaciones y Métodos*. New York: McGraw Hill

DeGroot, M. H. (1988). *Probabilidad y Estadística*. Iberoamericana: Addison- Wesley.

Devore, J. L. (2001). *Probabilidad y Estadística para ingeniería y ciencias* (5ª Edición). Mexico: Thomson Learning.

Evans, M. J. and Rosenthal, J. S. (2005). *Probabilidad y Estadística La Ciencia de la Incertidumbre*. Editorial Reverté.

Harnett and Murphy. (1993). *Introducción al Análisis Estadístico*. Iberoamericana: Addison-Wesley.

Montgomery and Runger. (1996). *Probabilidad y Estadística aplicadas a la Ingeniería*. New York: McGrawHill.

Walpole and Myers. (1992). *Probabilidad y Estadística* (4ª Edición). New York: McGraw Hill.