Journal of Kerbala University , Vol. 9 No.1 Scientific . 2011

Approximation Bayesian for selecting the least cell in multinomial population by functional analysis

التقريب البيزيني لاختيار الخلية الأصغر في مجتمع متعدد الحدود باستخدام التحليل الدالي

كوثر فوزي السلطاني
كلية التربية (أبن حيان)-قسم الرياضيات / جامعة بابل ,2010

Abstract:

Employment functional analysis to derive Bayesian approximation to select the smallest category (cell) in multinomial population, with linear loss function and prior Dirichlet distribution.

الملخص:

توظيف التحليل الدالي لاشتقاق التقريب البيزيني لاختيار (الخلية) الصنف الأصغر في مجتمع متعدد الحدود , بدالة خسارة خطية وتوزيع درشليت سابق .

1. Introduction

A ranking and selection problem usually results from questions like : "which brand of cigarettes is least likely to cause cancer ?" ; "which type of seat belt reduces car accident injuries the most ?";"which type of atomic power plant increases radiation levels the best?"; or" which type of catalyst increases a certain chemical process yield the most ?". Thus, we have several alternatives[KIM,S.H and NELSON,2003]

The general formulation which we use below is as follows. We have $k \geq 2$ sources of observation (each such source is called a population denoted by $\pi_1, \pi_2, \pi_1, ..., \pi_k$.From source i we may obtain independent and identically distributed observation $X_{i1}, X_{i2}, X_{i3}, X_{i4}, ...$ whose distribution involves an unknown parameter $\theta_i (1 \leq i \leq k)$; except for the value of θ_i the distribution is assumed not to differ from population to population . our goal is to select that population which has the largest parameter $\max(\theta_1, \theta_2, ..., \theta_k)$.we are to perform the selection in such a way that the probability $P(CS)$ of selecting the correct population is at least P^* (where P^* is a specified number between 1/k and 1)[SEONG,KIM and L.NELSON,2004]. Whenever the largest and next largest of $\theta_1, \theta_2, ..., \theta_k$ are "sufficiently far" apart (it being usually impossible to satisfy $P(CS) \geq P^*$ for all $\theta_1, \theta_2, ..., \theta_k$,since the θ's could then be arbitrarily close together); the demand that any proposed procedure satisfy this criterion is called the probability requirement. more general formulations can be considered (and will be noted in this paper) , but this one (with minor variations such as a goal of selecting that population which has the smallest parameter $\min(\theta_1, \theta_2, ..., \theta_k)$ accounts for more than 75 percent of the work in this area to date .[6,5]

2. Bayesian Multinomial Selection Problem[S.A.MADHI and K.F.HAMZA,2007]

Let $p_{[1]} \le p_{[2]} \le ... \le p_{[k]}$ denote the ordered values of the $p_i (1 \le i \le k)$ and let $\underline{n}$ has the multinomial distribution with probability mass function, such that the probability p_i of an observation in the cell i $P(\underline{n} \mid \underline{p}) = \frac{m!}{n_1! n_2! ... n_k!} \prod_{i=1}^{k} p_i^{n_i}$, $\sum_{i=1}^{k} n_i = m$ where $\underline{p} = (p_1, p_2, ..., p_k)$.

In the Bayesian procedure we depended prior and posterior distribution. Prior distribution of p_i is conjugate to the multinomial distribution Dirichlet distribution with prior density function is

$$\pi(\underline{p}) = \frac{\Gamma\left(\sum_{i=1}^{k} n_i'\right)}{\prod_{i=1}^{k} \Gamma(n_i')} \prod_{i=1}^{k} p_i^{n_i'-1} \qquad ....(1)$$

The distribution may be denoted by $Dr(n_1', n_2', ..., n_k', m')$, $m' = \sum_{i=1}^{k} n_i'$

and the posterior probability $\pi(\underline{p} \mid \underline{n}) \propto p_1^{n_1+n_1'-1} .... \ p_k^{n_k+n_k'-1}$

Since

$$P(\underline{n} \mid \underline{p}) \propto p_1^{n_1} .... \ p_k^{n_k} \qquad \text{and}$$

$$\pi(\underline{p}) \propto p_1^{n_1'-1} .... \ p_k^{n_k'-1}$$

This is a member of the Dirichlet family with parameters

$n_i'' = n_i' + n_i$ and $m'' = m' + m$ (i=1, 2, …., k) with mean $\hat{p}_i = \frac{n_i''}{m''}$

Will be termed the posterior frequency in the cell i[3].

3. Approximating of Bayesian Risk

The stopping risk (the posterior expected loss) of the terminal decision d_i when the posterior distribution for $\underline{P}$, that is when the sample path has reached $(n_1', n_2', ..., n_k', m')$ from the origin $(n_1', n_2', ..., n_k', m')$ denote by $S_i(n_1', n_2', ..., n_k', m')$ can be approximate by using theorem of functional analysis as follows:

$$S_i(n_1'', n_2'', ..., n_k''; m'') = \underset{\pi(\underline{p}|\underline{n})}{E}[L(d_i, \underline{p}^*)]$$

$$= mc + k^* \left[\frac{n_i''}{m''} - \underset{\pi(\underline{p}|\underline{n})}{E}(p_{[1]}) \right] \qquad ..... . (2)$$

Where k^* is the cost of sampling .

The derived of $\underset{\pi(\underline{p}|\underline{n})}{E}[\text{þ}_{[1]}]$ is

$$\underset{\pi(\underline{p}|\underline{n})}{E}[\text{þ}_{[1]}] = \int_0^1 p_{[1]} . g(p_{[1]}) dp_{[1]}$$

Such that, $g(p_{[1]}) = k\ f(p_{[1]}).\left[1 - F(p_{[1]})\right]^{k-1}$

And

$$f(p_{[1]}) = \frac{(m''-1)!}{(n''_{[1]}-1)!(m''-n''_{[1]}-1)!} p_{[1]}^{n''_{[1]}-1}(1-p_{[1]})^{m''-n''_{[1]}-1}$$

$$\underset{\pi(\underline{p}|\underline{n})}{E}(\text{þ}_{[1]}) = \frac{k(m''-1)!}{(n''_{[1]}-1)!(m''-n''_{[1]}-1)!}\left\{ \int_0^1 \left[1 - \left(1-p_{[1]}\right)^{m''-1}(m''-1)! \sum_{j_l=n''_{[1]}}^{m''-1} \frac{\left[\dfrac{p_{[1]}}{\left(1-p_{[1]}\right)}\right]^j}{j!(m''-1-j)!} \right]^{k-1} \right.$$

$$\left. p_{[1]}^{n''_{[1]}}(1-p_{[1]})^{m''-n''_{[1]}-1} \right\} dp_{[1]} \ . \qquad \text{..... . (3)}$$

Now, we can approximate the formula above by using the following theorem analysis

$(\mathit{if} \quad n \geq 1 \ \mathit{and} \quad x_i \in \Re \ \ \mathit{then} \ \left[x_1 + x_2\right]^n \leq 2^{n-1}\left[x_1{}^n + x_2{}^n\right], i = 1,2)$ [BHAYAA.E.S,2003]

Take the formula,

$$\left[1 - \sum_{j_l=n''_{[1]}}^{m''-1} \frac{(m''-1)!(1-p_{[1]})^{m''-1}\left[\dfrac{p_{[1]}}{\left(1-p_{[1]}\right)}\right]^j}{j!(m''-1-j)!} \right]^{k-1} = \sum_{J=0}^{k-1}\left\{ \binom{k-1}{J} (-1)^{Jl}((m''-1)!)^J((1-p_{[1]})^{m''-1})^J \right.$$

$$\left. \cdot \left(\sum_{j=n''_{[1]}}^{m''-1} \frac{\left(\dfrac{p_{[1]}}{1-p_{[1]}}\right)^j}{j!(m''-1-j)!} \right)^J \right\}$$

Put $J = l$, and using the above theorem after generalization we obtain, $$\left[\sum_{j_l=n''_{[1]}}^{m''-1}\frac{(m''-1)!\left[\frac{p_{[1]}}{\left(1-p_{[1]}\right)}\right]^{j}\left(1-p_{[1]}\right)^{(m''-1)}}{j!(m''-1-j)!}\right]^{l}$$

$$\leq 2^{(l-1)(m''-1)}\left\{\sum_{j=n''_{[1]}+1}^{m''-1}\frac{1}{2^{(l-1)(m''-1-j)}}\left(\frac{(m''-1)!\left[\frac{p_{[1]}}{\left(1-p_{[1]}\right)}\right]^{j-1}\left(1-p_{[1]}\right)^{(m''-1)}}{(j-1)!(m''-j)!}\right)^{l}+\left(\left[\frac{p_{[1]}}{\left(1-p_{[1]}\right)}\right]^{m''-1}\left(1-p_{[1]}\right)^{(m''-1)}\right)^{l}\right\}.$$

Then,

$$\underset{\pi(\underline{p}|\underline{n})}{E}(\text{þ}_{[1]})\leq\frac{k(m''-1)!}{(n''_{[1]}-1)!(m''-n''_{[1]}-1)!}\sum_{l=0}^{k-1}(-1)^{l}\binom{k-1}{l}2^{(l-1)(m''-1)}.$$

$$\left[\sum_{j_l=n''_{[1]}+1}^{m''-1}\left(\left(\frac{(m''-1)!}{(j-1)!(m''-j)!}\right)^{l}\cdot\frac{1}{2^{(l-1)(m''-1-j)}}\right)\left(\int_0^1\left[\frac{p_{[1]}}{\left(1-p_{[1]}\right)}\right]^{j-1}\left(\left(1-p_{[1]}\right)^{(m''-1)}\right)^{l}\left(p^{n''_{[1]}}{}_{[1]}\left(1-p_{[1]}\right)^{(m''-n''_{[1]}-1)}\right)dp_{n''_{[1]}}\right.\right.$$
$$\left.+\int_0^1\left(\left(p_{[1]}\right)^{(m''-1)}\right)^{l}\left(p^{n''_{[1]}}{}_{[1]}\left(1-p_{[1]}\right)^{(m''-n''_{[1]}-1)}\right)dp_{[1]}\right]\qquad\text{..... . (4)}$$

$$E(P_{[1]})\leq\frac{k(m''-1)!}{(n''_{[1]}-1)!(m''-n''_{[1]}-1)!}\sum_{l=0}^{k-1}.(-1)^{l}\binom{k-1}{l}2^{(l-1)(m''-1)}\left\{.\left(\sum_{j=n''_{[1]}+1}^{m''-1}\frac{\left((m''-1)!\big/(j-1)!(m''-1-j)!\right)^{l}}{2^{(l-1)(m''-1-j)}}\right).\right.$$
$$\left.\left\langle\frac{\Gamma\left(l(j-1)+n''_{[1]}+1\right)\ .\Gamma\left(l(m''-j)+m''-n''_{[1]}-1\right)}{.\Gamma\left(l(m''-j)+l(j-1)+m''\right)}\right\rangle+\frac{\Gamma\left(l(m''-1)+n''_{[1]}+1\right)\Gamma\left(m''-n''_{[1]}\right).}{\Gamma\left(l(m''-1)+m''+n''_{[1]}\right)}\right\}.$$

Then,

$$S_i(n''_1,n''_2,...,n''_k;m'')\leq\left\langle mc+k^{*}\left(\frac{n''_i}{m''}\right)-k^{*}\left(\frac{k(m''-1)!}{(n''_{[1]}-1)!(m''-n''_{[1]}-1)!}\sum_{l=0}^{k-1}(-1)^{l}\binom{k-1}{l}2^{(l-1)(m''-1)}.\right.\right.$$

$$\left\{\left(\sum_{j=n''_{[1]}+1}^{m''-1}\frac{\left((m''-1)!\big/(j-1)!(m''-1-j)!\right)^{l}}{2^{(l-1)(m''-1-j)}}\right)\left\langle\frac{\Gamma\left(l(j-1)+n''_{[1]}+1\right)\ .\Gamma\left(l(m''-j)+m''-n''_{[1]}-1\right)}{.\Gamma\left(l(m''-j)+l(j-1)+m''\right)}\right\rangle\right.$$

$$\left.\left.\left.+\frac{\Gamma\left(l(m''-1)+n''_{[1]}+1\right)\Gamma\left(m''-n''_{[1]}\right)}{\Gamma\left(l(m''-1)+m''+n''_{[1]}\right)}\right\}\right)\right\rangle.\qquad\text{..... . (5)}$$

4. Conclusion and Future Works

1- Conclusion:-

Ranking and selection procedures provides excellent tools for selecting the least of k competing alternatives. In this paper we apply Bayesian statistical decision theory which leads to a quite different approach to the selection problem as the concepts of loss of taking a certain decision when particular values of the parameters of interest are true, the cost of sampling and some prior information about the parameters of the underlying distributions are involved. It is quite clear that the problem of selecting the category with smallest probability is not equivalent (and not reducible) to that of selecting the category with the smallest probability[KIM,S.H and NELSON,2003] .in this paper we approximate the Bayesian posterior risk by using the functional analysis after the generalization to that theorem in functional analysis[BHAYAA.E.S,2003].

2- Future works:-

1. We can approximate by using the stirling's approximation
2. General loss functions may be used, where linear loss is considered as a special case.
3. Group sequential sampling can be tried where observations are taken in groups to build Bayesian sequential scheme for the selection problem. Since in this work we takes fixed sampling.
4. General loss functions may be tried, where linear loss is considered as a special case.
5. we can find the probability of correct selection in both cases (simulation and optimal) bayes fixed or group sampling selection procedure.

5. References

[1]- BHAYAA.E.S. On Constraint And Unconstraint Approximation ,PH,D.2003.

[2]- KIM, S.H., AND NELSON, B.L,. Selecting the best system. Proceedings 2003. Winter simulation conference, Georgia Institute of Technology, North Western University.

[3]- SAAD A.MADHI AND KAWTHER F .HAMZA ,Bayesian Fixed Sample Size Procedure for Selecting the Least Probable Event in Multinomal Distribution. Journal of college of education, university of Babylon. P391-369.number 36,2007, 1st Conference Scientific.

[4]- SEONG HEE, KIM GEORGIA and BARRY L. NELSON , Selecting the Best System ,Institute of Technology Northwestern University Proceedings of the 2004, Winter Simulation Conference.

[5]- JEAN D. GIBBONS,I.OLKIN AND M.SOBLE .Selecting And Ordering Population , 2009,Anew Statistical Methodology .

[6]- LARI ARJOMAND, Descriptive Statistics 1996,lecture three.