Università di Pisa
Dipartimento di Statistica e Matematica
Applicata all'Economia

Report n. 296

Strange phenomena in the most basic inferential procedure: interval estimation for a binominal proportion.

Emanuele Del Fava, Piero Manfredi

Pisa, luglio 2007
- Stampato in proprio -

Via Cosimo Ridolfi, 10 – 56124 PISA – Tel. Segr. Amm. 050 2216231 Segr. Stud. 050 2216317 Fax 050 2216375
Cod. Fisc. 80003670504 - P. IVA 00286820501 – Web http://statmat.ec.unipi.it- E-mail: dipstat@ec.unipi.it

Strange phenomena in the most basic inferential procedure: interval estimation for a binomial proportion

Emanuele Del Fava, Piero Manfredi *

Abstract

One of the most basic procedure in statistical inference is the construction of a confidence interval for an estimated proportion. Usually this task is performed by the so-called "standard interval" or "Wald interval", which is built adding and subtracting a certain quantity from the estimate, as it is taught in every elementary course of statistics.

However, the more recent statistical literature has revealed that the standard interval presents a serious problem: its actual coverage probability level is usually far from the nominal level, also when the estimated proportion is next to 0.5 or the sample size is large. This problem has also other consequences: indeed, a standard tool for the determination of the sample size n is based on the inversion of the Wald interval.

For these reasons, many authors have studied some alternative confidence intervals which try to solve, in different ways, the problem of the Wald interval.

1 Introduction

The estimation of a proportion is the most basic inferential problem. A main use of the estimated proportions is as the base for the determination of the sample size of a new survey.

For example, let us consider a serological survey, i.e. a survey aimed to estimate the seroprevalence (proportion of seropositive) for a certain infection, e.g. measles, in a given population by a sample of human blood specimens, usually stratified by age. See, for example, Farrington (1990) [7].

In some cases, for example, measles and varicella in the pre-vaccination era, seroprevalence ranges from values close to 0 in the youngest age groups to values close to 1 in the older age groups. Since the cost for every specimen of blood to be analysed can be high, it is necessary to determine a minimum sample size for every age groups [1]. For this reason we need a method that, given a fixed "nominal" coverage probability level and a fixed total size of error (usually chosen by the researcher), determines the minimum sample size. Usually the method consists in the inversion of a confidence interval. An example is the following well-known formula, from the "standard" or "Wald" interval:

$$n = \frac{4z^2_{\alpha/2}\hat{\pi}(1-\hat{\pi})}{A^2}, \tag{1}$$

*Dipartimento di Statistica e Matematica Applicata all'Economia, Facoltà di Economia, Università di Pisa, Via Cosimo Ridolfi 10 - 56124 Pisa, Italy.

[1]This determination is driven by the proportions estimated in previous surveys.

where A is the total size of error, i.e. the difference between the upper bound and the lower bound of the interval. In some textbooks this formula is accompanied by the caveat that the event should not be rare, e.g. $0.2 < \hat{\pi} < 0.8$. However, as we will see afterwards, this assumption is not sufficient, since the standard interval has a lot of problems also when $\hat{\pi}$ is next to 0.5.

1.1 The coverage probability of a confidence interval

The main criterion to judge the goodness of a confidence interval is its actual coverage probability level. Generally, when constructing an interval, we wish the actual coverage probability $C(\pi, n)$ be close to the nominal confidence level $1 - \alpha$. Because of the discrete nature of the binomial distribution, we can never achieve the exact nominal confidence level unless a randomized procedure is used. Thus our objective is to construct nonrandomized confidence intervals (CI) for π such that the coverage probability is:

$$C(\pi, n) = Pr(CI \text{ contains } \pi) \approx 1 - \alpha, \tag{2}$$

where α is the prespecified significance level.

As previously anticipated, the most used confidence interval for the proportion, the "standard interval" or "Wald interval", is not a good interval because it fails in almost every situation in reaching an actual coverage probability which is at least at the level $1 - \alpha$ (Brown *et al.*, 2001 [11]). For this reason, the recent literature has introduced some alternative confidence intervals that solve the problems of the Wald interval.

In this paper we first review the drawbacks of the standard interval as evidenced in the recent statistical literature, trying to a give a reason for them, and the corresponding remedies. Second, we show the importance of these achievements in the design of sample surveys with an illustration on seroprevalence surveys demonstrating that alternative solutions can provide much more reliable results in important problems. In Sec. 2 we show the problems of the Wald interval. In Sec. 3 we introduce the alternative confidence intervals. In Sec. 4 we report the illustration on seroprevalence surveys. Concluding remarks follow.

2 The Wald interval and its shortcomings

The Wald interval for the proportion π is based on a normal approximation and is obtained by inverting the acceptance region of the *Wald large-sample normal test* (Wald, 1943 [2]). The latter has the general form:

$$\text{Accept } H_0\text{: } \hat{\theta}\text{: } \left| \frac{\hat{\theta} - \theta}{\hat{se}(\hat{\theta})} \right| \leq z_{\alpha/2}, \tag{3}$$

where θ is a generic parameter, $\hat{\theta}$ is an estimate of θ and $\hat{se}(\hat{\theta})$ is the estimated standard error of $\hat{\theta}$. In the binomial case, if we want to test whether the estimate $\hat{\pi} = X/n$ (where X is the number of successes) is significantly equal to π, we have:

$$\text{Accept } H_0\text{: } \hat{\theta}\text{: } \left| \frac{\frac{X}{n} - \pi}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} \right| \leq z_{\alpha/2}. \tag{4}$$

By inverting 4, the standard CI is

$$CI_S = \hat{\pi} \pm z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}. \tag{5}$$

This CI is usually presented along with some justification based on the Central Limit Theorem.

However this CI has some serious problems: contrary to the common wisdom, its actual coverage probability $C(\pi, n)$ is often far from the nominal coverage probability level $1-\alpha$, even when n is large or π is far from 0 or 1. It happens that $C(\pi, n)$ of the Wald CI contains nonnegligible oscillation as both π and n vary. Brown *et al.* (2001) [11] show that there exist both "lucky" pairs (π, n) (such that $C(\pi, n)$ is very close to or larger than the nominal level $1-\alpha$) and "unlucky" ones (π, n) (such that $C(\pi, n)$ is much smaller than the nominal level).

The next examples show the inadequacy of the standard interval.

Fig. 1 plots the coverage probability of the nominal 95% standard interval for $\pi = 0.5$. The sample size n varies from 25 to 100. It is clear from the plot that the coverage probability does not steadily appoach the nominal confidence level as n increases and, moreover, the oscillation is significant. For instance, $C(0.2, 31) = 0.948$ and $C(0.2, 98) = 0.923$, so the coverage probability is significantly closer to 0.95 when $n = 31$ then when $n = 98$. It is noticeable that increasing the sample size n does not improve the actual coverage probability, as it occurs with inconsistent estimators!

Fig. 2 shows the coverage probability of the nominal 95% standard interval with fixed $n = 100$ and variable π from 0 to 1, with step 0.005. We see that, in spite of the "large" sample size, significant change in coverage probability occurs at the varying of π. Only for values belonging to the interval [0.2,0.8], the actual coverage probability level is close to the nominal 95%, even though it is often lower.

Anyway, we can observe from Fig. 2 that, about the question of the factors influencing mostly the convergence of the binomial to the normal distribution, it is more important the quickness to the convergence, which reaches its maximum at $\pi = 0.5$, than the small variance of the estimator.

The previous examples exhibit that the actual coverage probability of the Wald interval can differ significantly from the nominal confidence level for moderate and even large sample sizes and not only when π is near 0 or 1. There are two kinds of problems in the coverage probability of the standard interval: firstly, a systematic negative bias, that is the actual coverage probability level tends to be sistematically lower than the nominal level; secondly, an oscillatory behaviour for any values of n and π.

2.1 The reason for the bias

The bias arises from the use of a central interval for a discrete skewed distribution, meaning, among other things, that the standard interval has the "wrong" center (Brown *et al.*, 2001 [11]). The Wald interval is centered at $\hat{\pi} = X/n$. Although $\hat{\pi}$ is the maximum-likelihood estimate and an unbiased estimate of π, the choice of using it as the center of a confidence interval causes a systematic negative bias in the coverage. As we will see later on with the alternative confidence intervals, by suitably recentering the interval one can increase the coverage probability significantly for π away from 0 or 1 and almost always eliminate the bias.

We know, from the Central Limit Theorem, that

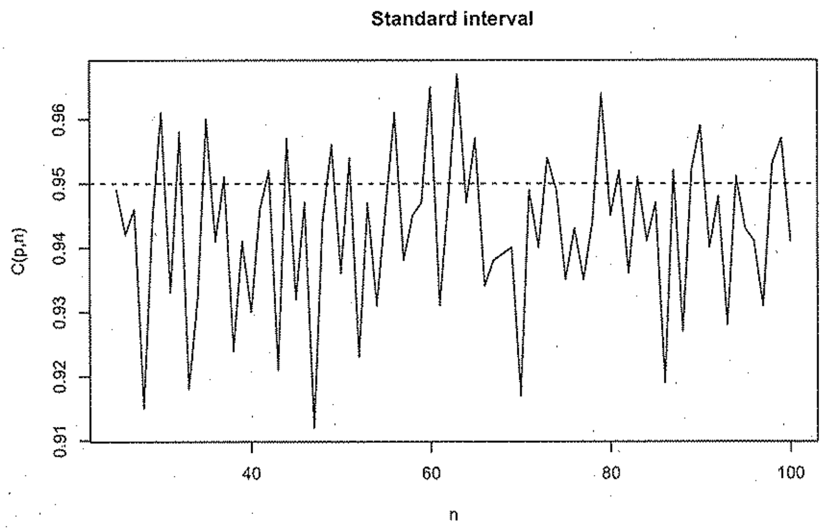

Figure 1: Standard interval: oscillations of the actual coverage probability level for fixed $\pi = 0.5$, variable sample size n=25 to 100 and nominal CP at 95% (dashed line).

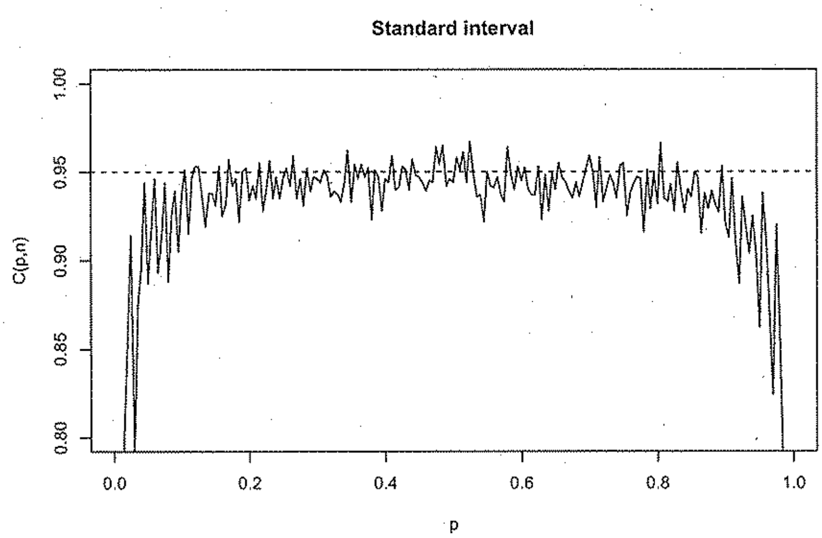

Figure 2: Standard interval: oscillations of the actual coverage probability level for fixed $n = 100$, variable π and nominal CP at 95% (dashed line).

$$Z_n = \frac{n^{1/2}(\hat{\pi} - \pi)}{\sqrt{\pi(1-\pi)}} \sim N(0,1). \tag{6}$$

The standard interval is based on the hypothesis that the Wald test W_n is asymptotically standard normally distributed:

$$W_n = \frac{n^{1/2}(\hat{\pi} - \pi)}{\sqrt{\hat{\pi}(1-\hat{\pi})}} \sim N(0,1). \tag{7}$$

The problem lies in the difference between Z_n and W_n: we assume that using $\hat{\pi}$ rather than π makes little difference, i.e. the distribution of W_n is the same of Z_n, whereas this is not all.

Let us compare with the case of the Wald test applied to the sampling mean $\bar{x}$. The variable Z_n is

$$Z_n = \frac{n^{1/2}(\bar{x} - \mu)}{\sigma} \sim N(0,1); \tag{8}$$

the respective Wald test W_n is:

$$W_n = \frac{n^{1/2}(\bar{x} - \mu)}{\sqrt{s^2}}, \tag{9}$$

where $s^2 = \sum_i (x_i - \bar{x})^2/(n-1)$ is the correct sampling variance. In this case, we know the distribution of W_n: it is t-distributed with $n-1$ degrees of freedom and approximates the Normal for $n > 20$.

By this comparison we can see that, what happens for the variance s, whose distribution is known, does not occur for $\hat{\pi}(1-\hat{\pi})$, whose distribution is unknown.

Instead, even for quite large values of n, the actual distribution of W_n in the binomial case is significantly nonnormal (Brown *et al.*, 2001 [11]). Thus, the very premise on which the standard interval is based is seriously compromised for moderate and even quite large values of n. In particular, for moderate n, however, the deviations of the bias, variance, skewness and kurtosis of W_n from their respective asymptotic values are often significant and cause a nonnegligible negative bias in the coverage probability of the standard confidence interval, so that the actual coverage $C(p,n)$ rarely reaches the nominal coverage $1-\alpha$.

We can analytically demonstrate the bias in the distribution of W_n. Let us write W_n in function of Z_n:

$$W_n(Z_n) = \frac{Z_n}{\sqrt{1 + (1-2\pi)Z_n/\sqrt{n\pi(1-\pi)} - Z_n^2/n}}. \tag{10}$$

A standard Taylor expansion and formulas for central moments of the binomial distribution yield the following approximation to the bias (Brown *et al.*, 2002, p. 166 [12]):

$$E[W_n(Z_n)] \approx \frac{\pi - 1/2}{\sqrt{n\pi(1-\pi)}} \left(1 + \frac{7}{2n} + \frac{9(\pi - 1/2)^2}{2n\pi(1-\pi)}\right). \tag{11}$$

It can be seen from Eq. 11 and from Fig. 3 that W_n has negative bias for $\pi < 0.5$, positive bias for $\pi > 0.5$ and no bias for $\pi = 0.5$. Also from the observation of the plot of the first derivative of $E[W_n]$ with respect to p, it is possible understand its behaviour.

From Fig. 4, we can see that the first derivative of $E[W_n]$ gets closer to 0 (without never reaching it) when p is near $1/2$.

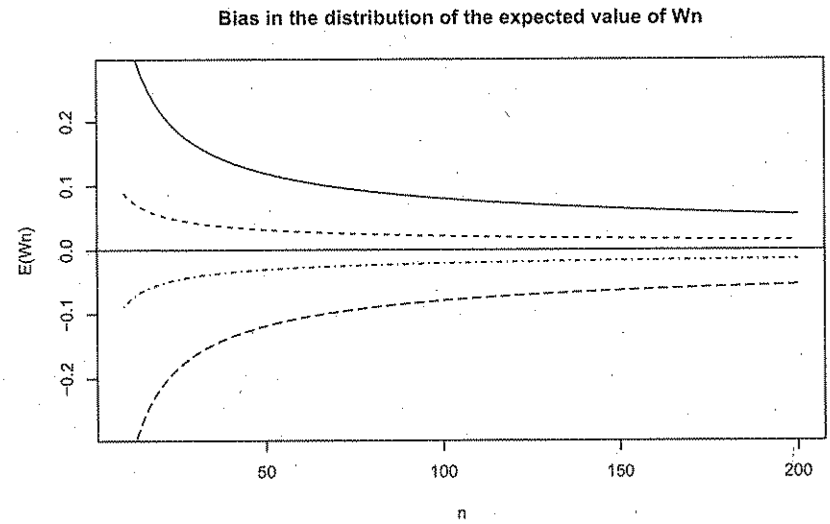

Figure 3: Wald interval: expected value of W_n for distint π values (starting from the top, $\pi_1 = 0.8, \pi_2 = 0.6, \pi_3 = 0.4, \pi_4 = 0.2$)

Therefore, temporary ignoring the oscillation effect, one can expect to increase the coverage probability by shifting the center of the standard interval towards 1/2, for which $E[W_n] = 0$.

2.2 The reason for the oscillation

In addition to the bias, the actual coverage probability of Wald interval oscillates in a significant way near the nominal coverage. The latter phenomenon comes from two sources: *discreteness* and *skewness* in the underlying binomial distribution. For a two-sided interval, the error due to discreteness is asymptotically dominant: it is of the order $1/\sqrt{n}$ and decreases when n increases (Brown *et al.*, 2002 [12]). On the contrary, the error due to skewness is secondary and is of the order $1/n$ (minor than $1/\sqrt{n}$), but still important fo even moderately large n. The oscillation in the coverage probability is therefore caused mainly by the discrete nature of the binomial distribution, more precisely by the interlaced structure of the binomial distribution.

Following again Brown *et al.* (2002) [12], let us try to understand at a more intuitive level why the coverage probability oscillates so significantly. By an easy calculation, one can show that the coverage probability $C(\pi, n) = Pr(\pi \in CI_s)$ equals $Pr(L_{\pi,n} \leq X \leq U_{\pi,n})$, where $L_{\pi,n}$ is the smallest integer larger than or equal to

$$\frac{n(\kappa^2 + 2n\pi) - \kappa n\sqrt{\kappa^2 + 4n\pi(1-\pi)}}{2(\kappa^2 + n)}, \tag{12}$$

and $U_{\pi,n}$ is the largest integer smaller than or equal to

$$\frac{n(\kappa^2 + 2n\pi) + \kappa n\sqrt{\kappa^2 + 4n\pi(1-\pi)}}{2(\kappa^2 + n)}. \tag{13}$$

What happens is that a small change in n or π can cause $L_{\pi,n}$ and/or $U_{\pi,n}$ to leap to the next integer value. For example, take the case $\pi = 0.5$ and $\alpha = 0.05$. When $n = 39$, we have

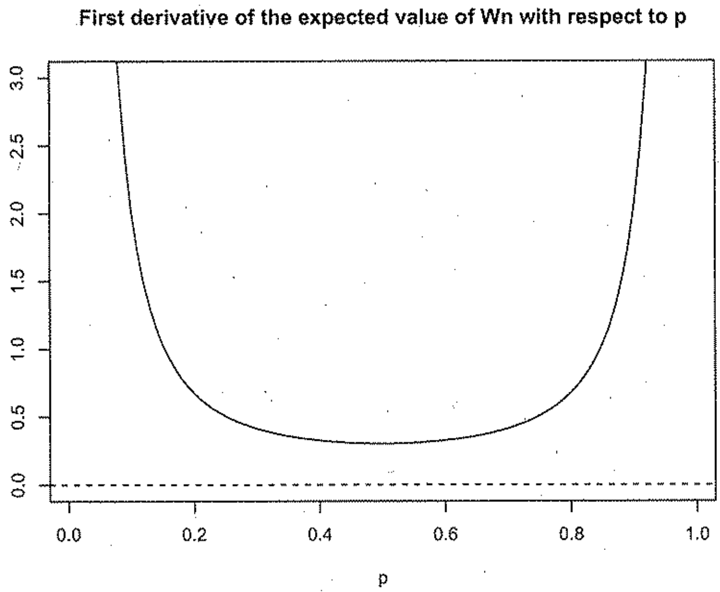

Figure 4: First derivative of $E[W_n]$ with respect to π

$L_{0.5,39} = 14$ and $U_{0.5,39} = 25$; but when $n = 40$, $L_{0.5,40}$ leaps to 15, while $U_{0.5,39}$ remains 25. Thus the set of favorable values of X loses the point $X = 14$ even though n has increased from 39 to 40. This causes $n = 40$ to be an unlucky choice of n: in effect, from data for Example 2, $C(0.5, 39) = 0.935$, while $C(0.5, 40) = 0.897$. This also happens when n is kept fixed and π changes slightly and so we begin to see unlucky values of π.

3 Recommended alternative intervals

The evidence of the shortcomings of the standard interval brings us to the consideration of alternative intervals. We now discuss several such alternatives presented in literature, most of all Brown *et al.* (2001) [11], each with its motivation.

3.1 The Wilson interval

The first alternative to the standard interval is obtained by simply getting back to Z_n (see 8) instead of W_n. The test is

$$\text{Accept } H_0\text{: } \hat{\theta}\text{: } Z_n = \left| \frac{\sqrt{n}(\hat{\pi} - \pi)}{\sqrt{\pi(1-\pi)}} \right| \leq \kappa, \tag{14}$$

and differs from Wald test because it uses the null standard error $(\pi(1 - \pi)/n)^{1/2}$ instead of the estimate standard error $(\hat{\pi}(1 - \hat{\pi})/n)^{1/2}$. This confidence interval can also be obtained by inverting the *Rao's tailed score test* of $H_0 : \mu = \mu_0$ (Rao, 1973 [9]). Indeed, score tests, and in particular their standard errors, are based on the log likelihood at the null hypothesis value of the parameter π, whereas Wald tests are based on the log likelihood at the maximum likelihood estimate (MLE) $\hat{\pi}$. If we solve, from 14, the ensuing quadratic equation $(\hat{\pi}-\pi)^2 = \kappa\pi(1-\pi)/n$ in terms of π, we have the confidence interval:

$$CI_W = \frac{X+\kappa^2/2}{n+\kappa^2} \pm \frac{\kappa\sqrt{n}}{n+\kappa^2}\sqrt{\hat{\pi}(1-\hat{\pi})+\kappa^2/4n}. \tag{15}$$

This interval was apparently introduced by Wilson (1927) [3] and so it is called the Wilson interval (or *score interval*, because it comes from the inversion of the score test).

The Wilson interval corrects the bias of the Wald interval, by recentering the interval at $\tilde{\pi} = (X+\kappa^2/2)/(n+\kappa^2)$, which is the weighted average of $\hat{\pi}$ and $1/2$, where n and κ^2 are the respective weights:

$$\tilde{\pi} = \frac{X+\kappa^2/2}{n+\kappa^2} = \frac{n\hat{\pi}+\kappa^2\frac{1}{2}}{n+\kappa^2}. \tag{16}$$

Therefore it falls between $\hat{\pi}$ and $1/2$, with the weight given to $\hat{\pi}$ approaching 1 aymptotically. This midpoint shrinks the sample proportion towards 0.5, the shrinking being less severe as n increases.

The coefficient of κ in the term that is added to and subtracted from the midpoint to form the score confidence interval can be rewritten in the following way:

$$\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\left[\frac{\sqrt{n}}{n+\kappa^2}\sqrt{n+\frac{\kappa^2}{4\hat{\pi}(1-\hat{\pi})}}\right]. \tag{17}$$

In this way, we can see that the error of Wilson is equal to the error of Wald multiplied for a certain factor, enclosed in brackets: if this factor is lower than 1, then the error of Wilson will be smaller than the Wald one.

Fig. 5 is the equivalent of Fig. 2 using Wilson instead of Wald interval. Not unsurprisingly the bias is basically removed since the source of error, i.e. the division by $\pi(1-\pi)$, is avoided, whereas the oscillation persists, since its origin, i.e. the discreteness, is not modifiable. The Wilson interval can be reccomended for use with nearly all sample sizes and parameter values. Coverage of this interval fluctuates acceptably near the nominal coverage $1-\alpha$, except for π very close to 0 or 1.

3.2 The Agresti-Coull interval

The standard interval CI_s is simple and easy to remember. For the purpose of classroom presentation and use in texts, it may be nice to have an alternative that has the familiar form $\hat{\pi} \pm z\sqrt{\hat{\pi}(1-\hat{\pi})/n}$, with a better and new choice of $\hat{\pi}$ rather than $\hat{\pi} = X/n$. Brown *et al.* (2001) [11] suggests that this can be accomplished by using the center of the Wilson region in place of $\hat{\pi}$. Given the following notation, $\widetilde{X} = X+\kappa^2/2$, $\tilde{n} = n+\kappa^2$, $\tilde{\pi} = \widetilde{X}/\tilde{n}$ and $\tilde{q} = 1-\tilde{\pi}$, we define the Agresti-Coull interval for π by

$$CI_{AC} = \tilde{\pi} \pm \kappa\sqrt{\tilde{\pi}\tilde{q}/\tilde{n}}. \tag{18}$$

Both the Agresti-Coull and the Wilson interval are centered on the same value, $\tilde{\pi}$. It is easy to check that the Agresti-Coull interval is never shorter than the Wilson one. For the case when $\alpha = 0.05$, if we use the value 2 instead of 1.96 for κ, this interval is the "add 2 successes and 2 failures" interval in Agresti and Coull (1998) [1]. For this reason, we call it the Agresti-Coull interval.

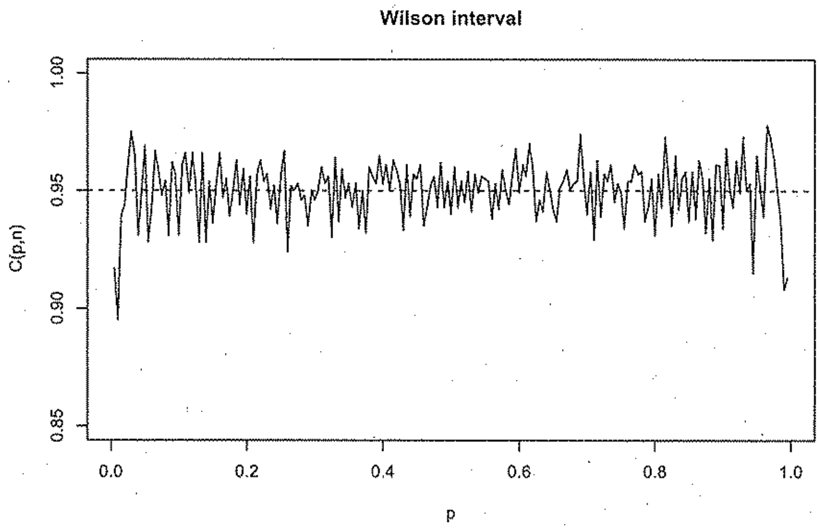

Figure 5: Wilson interval: coverage of the nominal 95% standard interval for fixed $n = 100$ and variable π

The Agresti-Coull interval has good minimum coverage probability. The coverage probability of the interval is quite conservative for π very close to 0 or 1. In comparison to the Wilson interval it is more conservative, especially for small n. However, it is to be noticed that this follows from the larger amplitude of Agresti-Coull. Thus, it is evident that exists a trade-off between the conservativeness and the length of an interval: from the one hand, a more conservative interval tends to be vaguer; on the other hand, a shorter interval can have a lower actual coverage probability. See Fig. 6.

3.3 The Jeffreys prior interval

The Jeffreys prior interval is a Bayesian interval based on the continuous Beta distribution, since the Beta is the standard conjugate prior distribution of the binomial.

Let us start with a Bayesian confidence interval. Suppose $X \sim Bin(n, \pi)$ and suppose π has a prior distribution $Beta(a_1, a_2)$; then the posterior distribution of π is $Beta(X + a_1, n - X + a_2)$. Thus a $100(1-\alpha)\%$ equal-tailed Bayesian interval is given by

$$L = \left[B\left(\frac{\alpha}{2}; X + a_1, n - X + a_2\right)\right] \tag{19}$$

and

$$U = \left[B\left(1 - \frac{\alpha}{2}; X + a_1, n - X + a_2\right)\right], \tag{20}$$

where $B(\alpha; m_1, m_2)$ denotes the α quantile of a $Beta(m_1, m_2)$ distribution.

The Jeffreys prior interval is a special case of a Bayesian interval and its distribution is $Beta(1/2, 1/2)$; for references see Berger (1985) [4]. Therefore the $100(1-\alpha)\%$ equal-tailed Jeffreys prior interval is defined as

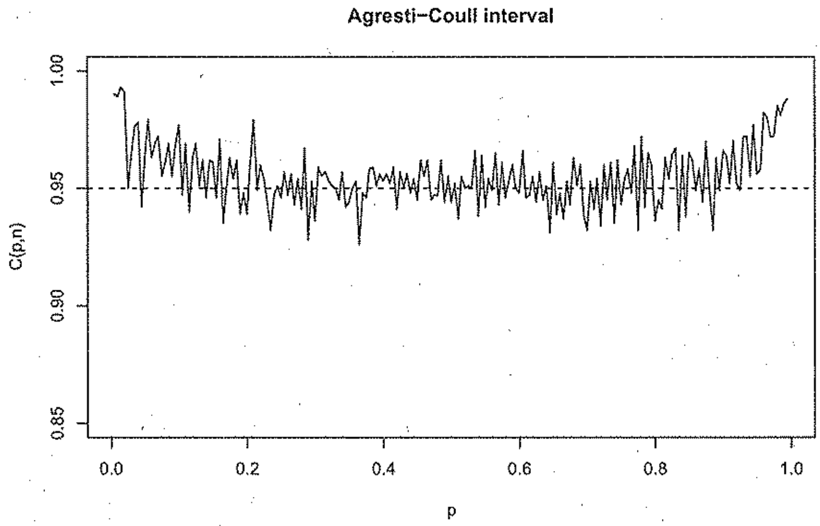

Figure 6: Agresti-Coull interval: coverage of the nominal 95% standard interval for fixed $n = 100$ and variable π

$$CI_J = [L_j(x), U_j(x)], \tag{21}$$

where $L_j(0) = 0$ and $U_J(n) = 1$ and otherwise

$$L_J(x) = \left[B_{\alpha/2}\left(X + \frac{1}{2}, n - X + \frac{1}{2}\right)\right], \tag{22}$$

$$U_J(x) = \left[B_{1-\alpha/2}\left(X + \frac{1}{2}, n - X + \frac{1}{2}\right)\right]. \tag{23}$$

The endpoints of the Jeffreys prior interval are the $\alpha/2$ and $1-\alpha/2$ quintiles of the $Beta(x + 1/2, n - x + 1/2)$ distribution.

The quality of the Jeffreys prior interval is qualitatively similar to that of CI_W over most of the parameter space [0,1]. The coverage has an unfortunate fairly deep spike near $\pi = 0$ and, simmetrically, another near $\pi = 1$. See Fig. 7.

3.3.1 The Clopper-Pearson interval

The Clopper-Pearson "exact" confidence interval, proposed for the first time by Clopper and Pearson (1934) [5], is based on the inversion of the equal-tailed binomial test of $H_0 : \pi = \pi_0$, rather than its normal approximation. It has endpoints that are the solutions in π_0 to the equations

$$\sum_{k=x}^{n} \binom{n}{k} \pi_0^k (1 - \pi_0)^{n-k} = \frac{\alpha}{2} \tag{24}$$

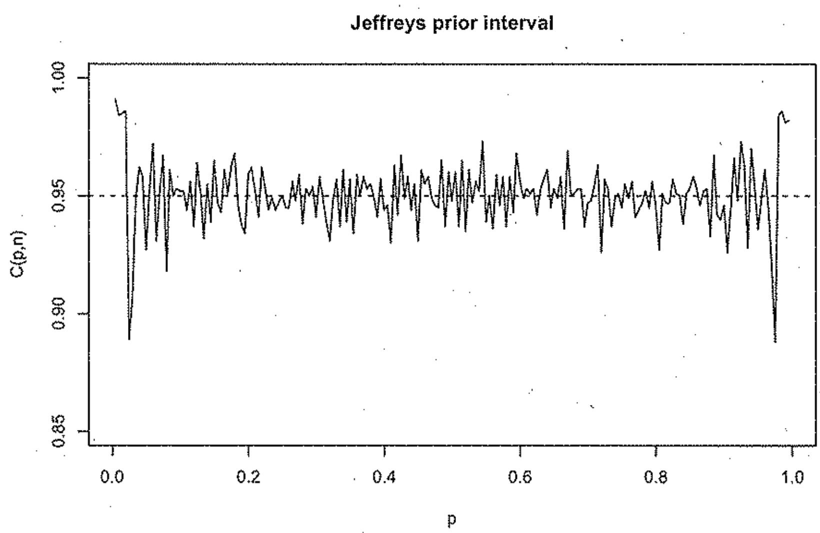

Figure 7: Jeffreys prior interval: coverage of the nominal 95% standard interval for fixed $n = 100$ and variable π

and

$$\sum_{k=0}^{x} \binom{n}{k} \pi_0^k (1-\pi_0)^{n-k} = \frac{\alpha}{2}, \tag{25}$$

except that the lower bound is 0 when $x = 0$ and the upper bound is 1 when $x = n$. This interval estimator is guaranteed to have coverage probability of at least $1 - \alpha$ for every possible value of π. When $x = 1, 2, \ldots, n-1$, the confidence interval equals

$$\left[1 + \frac{n-x+1}{xF_{2x,2(n-x+1),1-\alpha/2}}\right]^{-1} < \pi < \left[1 + \frac{n-x}{(x+1)F_{2(x+1),2(n-x),\alpha/2}}\right]^{-1}, \tag{26}$$

where $F_{a,b,c}$ denotes the $1-c$ quantile from the F distribution with degrees of freedom a and b. Equivalently, the lower endpoint is the $\alpha/2$ quantile of a beta distribution $Beta(x, n-x+1)$ and the upper bound is the $1-\alpha/2$ quantile of a beta distribution $Beta(x+1, n-x)$.

The Clopper-Pearson exact interval is typically treated as the "gold standard", although this procedure is necessarily very conservative, because of the discreteness of the binomial distribution. For any fixed parameter value, the actual coverage probability can be much larger than the nominal confidence level unless n is quite large. See Fig. 8.

3.4 Coverage probability

To judge the quality of these alternative confidence intervals, let us consider their average coverage probability, the average being over π. There is striking difference in the average coverage probability among the five intervals previously introduced. The standard interval performs poorly. The Clopper-Pearson "exact" interval is the more conservative, overall with the smaller values of n. The interval CI_{AC} is slightly conservative, but less than the Clopper-Pearson.

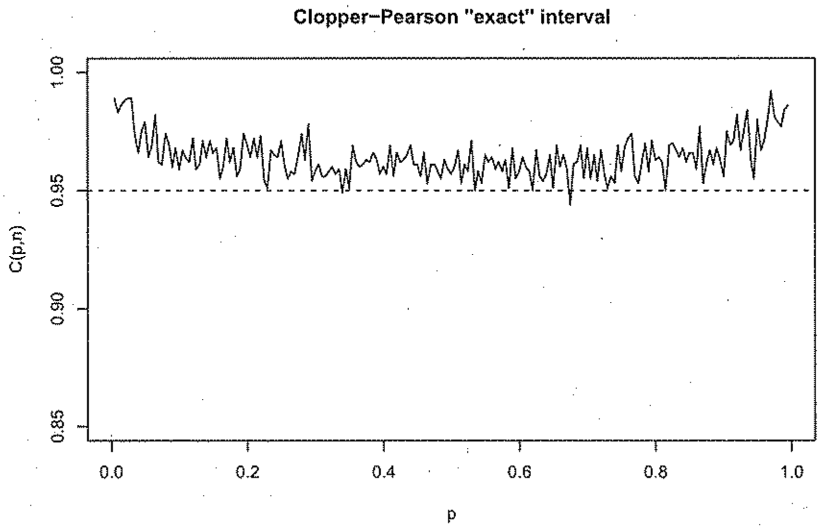

Figure 8: Clopper-Pearson "exact" interval: coverage of the nominal 95% standard interval for fixed $n = 100$ and variable π

Both the Wilson interval and the Jeffreys prior interval have excellent performances; that of the Jeffreys prior interval is, if anything, slightly superior. The average coverage probability of the Jeffreys interval is really very close to the nominal level even for quite small n.

3.5 Lengths of the intervals

The length is also an important feature of a good confidence interval, the actual coverage probability being equal. In general, the shorter the interval, the better it is. However, sometimes a longer interval can be preferable to a shorter one, e.g. because this interval allows a higher coverage probability level.

The shortest confidence interval is the Jeffreys prior interval when $0.10 < \hat{\pi} < 0.25$ and when $0.75 < \hat{\pi} < 0.99$, that is when $\hat{\pi}$ is near the boundaries (but not when $\hat{\pi} = 0$ or 1). Instead, when $0.25 < \hat{\pi} < 0.75$ the shortest interval are the Wilson and the Agresti-Coull, although the Wilson is sometimes more parsimonious than the Agresti-Coull. On the contrary, the longest confidence interval is the Clopper-Pearson "exact" interval, which is the more conservative one: in effect this is characterized by its length, larger than that of the other intervals, and by its actual coverage probability, which is very often over the nominal level. Finally, as regards the standard interval, it can happen that the upper bound is higher than 1, when the estimated proportion is next to 1.

4 Application of the alternative confidence intervals

We present now an application of the standard interval and the other alternative confidence intervals to varicella-zoster virus (VZV) data, from Thiry *et al.* (2002) [6]. Starting from the estimated proportions of that article, we will try to determine the optimal sample size for every age class using the intervals previously introduced. These data are presented in Tab. 1 for people aged 1-10 years old.

In general, when the confidence interval is given by a formula (and not by the quantiles of a continuous distribution), the optimal sample size can be calculated inverting it and looking for that value of n which allows a certain value of the total size of error, chosen *a priori* by the researcher. This total size of error, $A(n, \hat{\pi}, \kappa)$ is given by the difference between the upper bound U and the lower bound L of the interval and is function of the sample size n, the estimated proportion π and the quantile κ. For example, in case of the standard interval we have:

$$\begin{aligned} A_s(n, \hat{\pi}, \kappa) &= U_s - L_s \\ &= \hat{\pi} + \kappa\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} - \left(\hat{\pi} - \kappa\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right) \\ &= 2\kappa\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}. \end{aligned} \quad (27)$$

Also the value of $\hat{\pi}$ is chosen by the researcher: if he has some information about the true value of the proportion in the population we are studying, e.g. from previous studies, then he could use that value in order to calculate the optimal sample size; otherwise, unless he has any information about the true value of the parameters, then it is possible to use the value which maximizes the variance and so maximizes the sample size. Being the proportion π binomial distributed, the variance of X/n is maximized when $\pi = 0.5$.

As regards the Wilson interval or the Agresti-Coull interval, which are calculated without referring to a continuous distribution, the sample size n can be simply determined in the same way used for the standard interval (see Eq. 27). Since the formulae for these two intervals are more complex than that of the Wald CI, from the difference $(U - L)$ we obtain polynomials in n of order superior to 1. Anyway, the roots of these polynomials can be simply computed numerically with a software for numerical calculation (in this paper we have used the statistical software R 2.5.0 [10] for the data analysis). For the choice of the right root, we have to discart the complex roots and, if there more real roots, take the largest positive one.

The second-order polynomial equation in n relative to Wilson interval is

$$A^2n^2 + 2\kappa^2[A^2 - 2\hat{\pi}(1-\hat{\pi})]n + \kappa^4(A^2 - 1) = 0, \quad (28)$$

where $\kappa = z_{\alpha/2}$.

As regards the Agresti-Coull interval, we obtain the following third-order polynomial equation in n:

$$A^2n^3 + \kappa^2[3A^2 - 4\hat{\pi}(1-\hat{\pi})]n^2 + \kappa^4(3A^2 - 2)n + \kappa^6(A^2 - 1) = 0. \quad (29)$$

Both the equations 28 and 29 are "well-behaved", i.e. they admit always one and only one real positive root, which can represent the minimum sample size we are looking for. This follows from the analysis of the coefficients' signs of the equations in accordance to Cartesius' sign rule.

In addition to prevalence proportions, Tab. 1 presents the sample size n determined by the inversion of Wald, Wilson and Agresti-Coull intervals for A set *a priori* to 0.10. To judge these results, we have to remind what we have said about the length of the intervals, that is if an interval is longer, its inversion gives a larger sample size n. Therefore, the data presented in the table reflect this situation. When $0.25 < \hat{\pi} < 0.75$, which usually happens in the younger age

Age	Prevalence	n from Wald	n from Wilson	n from Agresti-Coull
1	0.3529	351	347	348
2	0.4737	383	379	379
3	0.5714	376	373	373
4	0.6786	335	332	332
5	0.8019	244	242	245
6	0.8478	198	198	201
7	0.8235	223	222	225
8	0.8830	159	160	165
9	0.9252	106	112	120
10	0.9451	80	89	98

Table 1: Seroprevalence for VZV serum IgG antibodies according to age and the sample size n determined from Wald, Wilson and Agresti-Coull intervals for $A = 0.10$

groups, both Wilson and Agresti-Coull give the most satisfying results (but Wilson is slightly better). On the contrary, when $\hat{\pi} < 0.25$ or $\hat{\pi} > 0.75$, it may seem that Wald gives the minor values: however, these values are not trustable because of the poor actual coverage probability of this interval.

Differently, when the interval is based on the quantiles of a continuous distribution, such as the Jeffreys prior interval and the Clopper-Pearson interval calculated from the *Beta*, the best way for the determination of n is to build the curves of the function $A(n, \hat{\pi}, \kappa)$: these curves can give and idea of how much the sample size should be large. Fig. 9 shows the curves from Jeffreys prior interval and Fig. 10 the curves from the Clopper-Pearson. From the observation of these graphs, it is plain that the Clopper-Pearson determines values of n averagely larger because of its conservativeness. Then, if we think to the length of Jeffreys prior interval, it might be that for $0.10 < \hat{\pi} < 0.25$ and $0.75 < \hat{\pi} < 0.99$ it gives smaller sample sizes, minor than Wald interval.

5 Conclusions

In this paper we have shown that the standard confidence interval for a proportion, one of the most commom statistical tools, presents heavy bias that compromise its efficacy noticeably: we remark here that it does not work in almost every situation, also for large values of the sample size or for $\pi = 0.5$, when the binomial distribution is symmetrical and converges to the Normal more rapidly. Nevertheless, the deep reasons of this strange behaviour are still not very clear and thus other studies are necessary.

Because of its problems, Wald interval should not be used even for the determination of the sample size, mainly for two reasons: firstly, if the interval does not run well, it is incorrect using it in any case, for it may take to systematic errors; secondly, there are other intervals that run better and effectively give smaller sample sizes.

The literature reports lots of alternative confidence intervals, not only those introduced in this paper: some are simpler to use, other less, but anyway they are correct and their actual coverage probability reaches truly the nominal level. Besides, nowadays the last statistical softwares, such as R 2.5.0 [10], allow to compute them in a feasible way, so it ought to be discouraged the use of Wald interval and promoted the use of these other intervals.

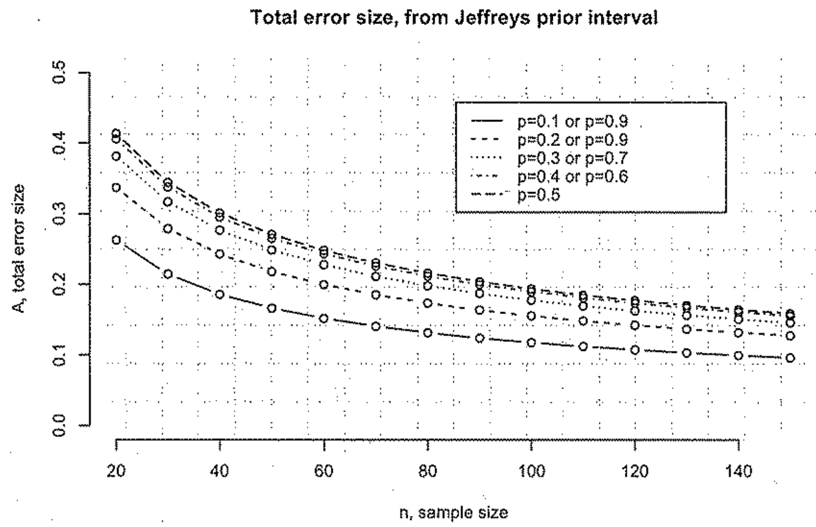

Figure 9: Curves of the function $A(n, \pi)$ from the Jeffreys prior interval

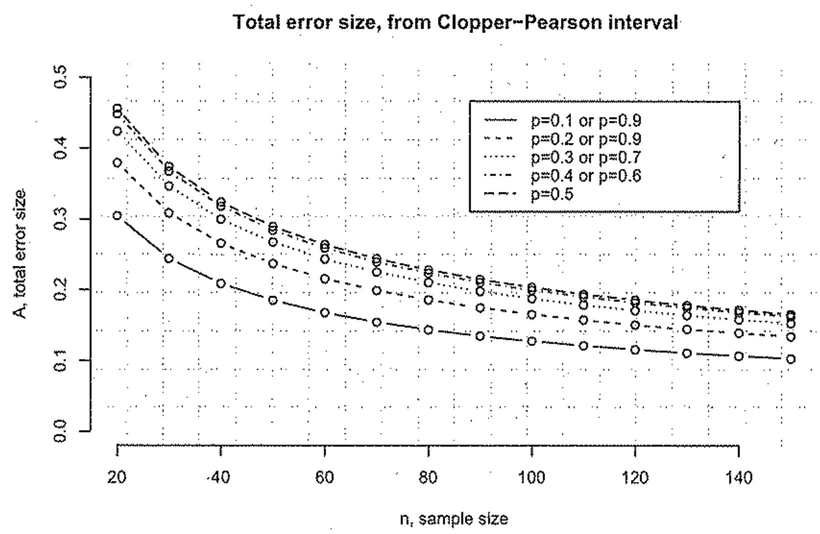

Figure 10: Curves of the function $A(n, \pi)$ from the Clopper-Pearson interval

References

[1] Agresti A. and Coull B. A. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, 52(2):119–126, 1998.

[2] Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3):426–482, 1943.

[3] Wilson E. B. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927.

[4] Berger J. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 1985.

[5] Clopper C. J. and Pearson E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26:404–413, 1934.

[6] Thiry N. Beutels P. Shkedy Z. Vranckx R. Vandermeulen C. Van Der Wielen M. and Van Damme P. The seroepidemiology of primary varicella-zoster virus infection in flanders (belgium). *European Journal of Pediatrics*, 161:588–593, 2002.

[7] Farrington C. P. Modelling forces of infection for measles, mumps and rubella. *Statistics in Medicine*, 9:953–967, 1990.

[8] Blyth C. R. and Still H. A. Binomial confidence intervals. *Journal of the American Statistical Association*, 78(381):108–116, 1983.

[9] Rao C. R. *Linear statistical inference and its applications*. Wiley, New York, 1973.

[10] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2007. ISBN 3-900051-07-0.

[11] Brown L. D. Cay T. T. and DasGupta A. Interval estimation for a binomial proportion (with discussion). *Statistical Science*, 16(2):101–133, 2001.

[12] Brown L. D. Cay T. T. and DasGupta A. Confidence intervals for a binomial proportion and asymptotic expansions. *The Annals of Statistics*, 30(1):160–201, 2002.

[13] Brown L. D. Cay T. T. and DasGupta A. Interval estimation in exponential families. *Statistica Sinica*, 13:14–49, 2003.

Elenco dei report pubblicati

Anno: **1987**

- ☐ **n. 1** Alberto Cambini - Laura Martein, Some Optimality Conditions in Vector Optimization
- ☐ **n. 2** Alberto Cambini - Laura Martein - S.Schaibel, On Maximizing a Sum of Ratios
- ☐ **n. 3** Giuliano Gasparotto, On the Charnes-Cooper Transformation in linear Fractional Programming.
- ☐ **n. 4** Alberto Cambini, Non-linear separation Theorems, Duality and Optimality
- ☐ **n. 5** Giovanni Boletto, Indicizzazione parziale: aspetti metodologici e riflessi economici
- ☐ **n. 6** Alberto Cambini - Claudio Sodini, On Parametric Linear Fractional Programming
- ☐ **n. 7** Alberto Bonaguidi, Alcuni aspetti meno noti delle migrazioni in Italia
- ☐ **n. 8** Laura Martein - S. Schaible, On Solving a Linear Program with one Quadratic Constraint

Anno: **1988**

- ☐ **n. 9** Ester Lari, Alcune osservazioni sull'equazione funzionale Ø(x,y,z)=Ø(Ø(x,y,t),t,z)
- ☐ **n. 10** F. Bartiaux, Une étude par ménage des migrations des personnes âgées: comparaison des résulttats pour l'Italie et les Etats-Unis
- ☐ **n. 11** Giovanni Boletto, Metodi di scomposizione del tasso di inflazione
- ☐ **n. 12** Claudio Sodini, A New Algorithm for the Strictly Convex Quadratic Programming Problem
- ☐ **n. 13** Laura Martein, On Generating the Set of all Efficient Points of a Bicriteria Fractional Problem
- ☐ **n. 14** Laura Martein, Applicazioni della programmazione frazionaria nel campo economico-finanziario
- ☐ **n. 15** Laura Martein, On the Bicriteria Maximization Problem
- ☐ **n. 16** Paolo Manca, Un prototipo di sistema esperto per la consulenza finanziaria rivolta ai piccoli risparmiatori
- ☐ **n. 17** Paolo Manca, Operazioni Finanziarie di Soper e Operazioni di puro Investimento secondo Teichroew-Robichek-Montalbano
- ☐ **n. 18** Paolo Carraresi - Claudio Sodini, A k - Shortest Path Approach to the Minimum Cost Matching Problem.
- ☐ **n. 19** Odo Barsotti - Marco Bottai, Sistemi gravitazionali e fasi di transazione della crescita Demografica
- ☐ **n. 20** Giovanni Boletto, Metodi di scomposizione dell'inflazione aggregata : recenti sviluppi.
- ☐ **n. 21** Marc Termote - Alberto Bonaguidi, Multiregional Stable Populatioin as a Tool for Short-term Demographic Analysis
- ☐ **n. 22** Marco Bottai, Storie familiari e storie migratorie: un'indagine in Italia
- ☐ **n. 23** Maria Francesca Romano - Marco Marchi, Problemi connessi con la disomogeneità dei gruppi sottoposti a sorveglianza statistico-epidemiologica.
- ☐ **n. 24** Franca Orsi, Un approcio logico ai problemi di scelta finanziaria.

Anno: **1989**

- ☐ **n. 25** Vincenzo Bruno, Attrazione ed entropia.
- ☐ **n. 26** Giorgio Giorgi - S. Mititelu, Invexity in nonsmooth Programming.
- ☐ **n. 28** Alberto Cambini - Laura Martein, Equivalence in linear fractional programming.

Anno: **1990**

- ☐ **n. 27** Vincenzo Bruno, Lineamenti econometrici dell'evoluzione del reddito nazionale in relazione ad altri fenomeni economici
- ☐ **n. 29** Odo Barsotti - Marco Bottai - Marco Costa, Centralità e potenziale demografico per l'analisi dei comportamenti demografici: il caso della Toscana
- ☐ **n. 30** Anna Marchi, A sequential method for a bicriteria problem arising in portfolio selectiontheory.
- ☐ **n. 31** Marco Bottai, Mobilità locale e pianificazione territoriale.
- ☐ **n. 32** Anna Marchi, Solving a quadratic fractional program by means of a complementarity approach
- ☐ **n. 33** Anna Marchi, Sulla relazione tra un problema bicriteria e un problema frazionario.

Anno: **1991**

- ☐ **n. 34** Enrico Gori, Variabili latenti e "self-selection" nella valutazione dei processi formativi.
- ☐ **n. 35** Piero Manfredi - E. Salinelli, About an interactive model for sexual Populations.
- ☐ **n. 36** Giorgio Giorgi, Alcuni aspetti matematici del modello di sraffa a produzione semplice
- ☐ **n. 37** ALberto Cambini - S.Schaibl - Claudio Sodini, Parametric linear fractional programming for an unbounded feasible Region.
- ☐ **n. 38** I.Emke - Poulopoulos - V.Gozàlves Pèrez - Odo Barsotti - Laura Lecchini, International migration to northern Mediterranean countries the cases of Greece, Spain and Italy.
- ☐ **n. 39** Giuliano Gasparotto, A LP code implementation
- ☐ **n. 40** Riccardo Cambini, Un problema di programmazione quadratica nella costituzione di capitale.
- ☐ **n. 41** Gilberto Ghilardi, Stime ed errori campionari nell'indagine ISTAT sulle forze di lavoro.
- ☐ **n. 42** Vincenzo Bruno, Alcuni valori medi, variabilità paretiana ed entropia.
- ☐ **n. 43** Giovanni Boletto, Gli effetti del trascinamento dei prezzi sulle misure dell'inflazione: aspetti metodologici
- ☐ **n. 44** P. Paolicchi, Gli abbandoni nell'università: modelli interpretativi.
- ☐ **n. 45** Maria Francesca Romano, Da un archivio amministrativo a un archivio statistico: una proposta metodologica per i dati degli studenti universitari.
- ☐ **n. 46** Maria Francesca Romano, Criteri di scelta delle variabili nei modelli MDS: un'applicazione sulla popolazione studentesca di Pisa.
- ☐ **n. 47** Odo Barsotti - Laura Lecchini, Les parcours migratoires en fonction de la nationalité. Le cas de l'Italie.
- ☐ **n. 48** Vincenzo Bruno, Indicatori statistici ed evoluzione demografica, economica e sociale delle province toscane.
- ☐ **n. 49** Alberto Cambini - Laura Martein, Tangent cones in optimization.
- ☐ **n. 50** Alberto Cambini - Laura Martein, Optimality conditions in vector and scalar optimization: a unified approach.

Anno: **1992**

- ☐ **n. 51** Gilberto Ghilardi, Elementi di uno schema di campionamento areale per alcune rilevazioni ufficiali in Italia.
- ☐ **n. 52** Paolo Manca, Investimenti e finanziamenti generalizzati.
- ☐ **n. 53** Laura Lecchini - Odo Barsotti, Le rôle des immigrés extra- communautaires dans le marché du travail

- ☐ n. 54 Riccardo Cambini, Alcune condizioni di ottimalità relative ad un insieme stellato.
- ☐ n. 55 Gilberto Ghilardi, Uno schema di campionamento areale per le rilevazionninsulle famiglie in Italia.
- ☐ n. 56 Riccardo Cambini, Studio di una classe di problemi non lineari: un metodo sequenziale.
- ☐ n. 57 Riccardo Cambini, Una nota sulle possibili estensioni a funzioni vettoriali di significative classi di funzioni concavo-generalizzate.
- ☐ n. 58 Alberto Bonaguidi - Valerio Terra Abrami, Metropolitan aging transition and metropolitan redistribution of the elderly in Italy.
- ☐ n. 59 Odo Barsotti - Laura Lecchini, A comparison of male and female migration strategies: the cases of African and Filipino Migrants to Italy.
- ☐ n. 60 Gilberto Ghilardi, Un modello logit per lo studio del fenomeno delle nuove imprese.
- ☐ n. 61 S. Schaible, Generalized monotonicity.
- ☐ n. 62 Vincenzo Bruno, Dell'elasticità in economia e dell'incertezza statistica.
- ☐ n. 63 Laura Martein, Alcune classi di funzioni concave generalizzate nell'ottimizzazione vettoriale
- ☐ n. 64 Anna Marchi, On the relationships between bicriteria problems and non-linear programming problems.
- ☐ n. 65 Giovanni Boletto, Considerazioni metodologiche sul concetto di elasticità prefissata.
- ☐ n. 66 Laura Martein, Soluzione efficienti e condizioni di ottimalità nell'otti-mizzazione vettoriale.

Anno: **1993**

- ☐ n. 67 Maria Francesca Romano, Le rilevazioni ufficiali ISTAT della popolazione universitaria: problemi e definizioni alternative.
- ☐ n. 68 Marco Bottai - Odo Barsotti, La ricerca "Spazio Utilizzato" Obiettivi e primi risultati.
- ☐ n. 69 Marco Bottai - F.Bartiaux, Composizione familiare e mobilità delle persone anziane. Una analisi regionale.
- ☐ n. 70 Anna Marchi - Claudio Sodini, An algorithm for a non-differentiable non -linear fractional programming problem.
- ☐ n. 71 Claudio Sodini - S.Schaible, An finite algorithm for generalized linear multiplicative programming.
- ☐ n. 72 Alberto Cambini - Laura Martein, An approach to optimality conditions in vector and scalar optimization.
- ☐ n. 73 Alberto Cambini - Laura Martein, Generalized concavity and optimality conditions in vector and scalar optimization.
- ☐ n. 74 Riccardo Cambini, Alcune nuove classi di funzioni concavo-generalizzate.

Anno: **1994**

- ☐ n. 75 Alberto Cambini - Anna Marchi - Laura Martein, On nonlinear scalarization in vector optimization.
- ☐ n. 76 Maria Francesca Romano - Giovanna Nencioni, Analisi delle carriere degli studenti immatricolati dal 1980 al 1982.
- ☐ n. 77 Gilberto Ghilardi, Indici statistici della congiuntura.
- ☐ n. 78 Riccardo Cambini, Condizioni di efficienza locale nella ottimizzazione vettoriale.
- ☐ n. 79 Odo Barsotti - Marco Bottai, Funzioni di utilizzazione dello spazio.
- ☐ n. 80 Vincenzo Bruno, Alcuni aspetti dinamici della popolazione dei comuni della Toscana, distinti per ampiezza demografica e per classi di urbanità e di ruralità.
- ☐ n. 81 Giovanni Boletto, I numeri indici del potere d'acquisto della moneta.
- ☐ n. 82 Alberto Cambini - Laura Martein - Riccardo Cambini, Some optimality conditions in multiobjective programming.
- ☐ n. 83 S. Schaible, Fractional programming with sume of ratios.
- ☐ n. 84 Stefan Tigan - I.M.Stancu-Minasian, The minimun-risk approach for continuous time linear-fractional programming.
- ☐ n. 85 Vasile Preda - I.M.Stancu-Minasian, On duality for multiobjective mathematical programming of n-set.
- ☐ n. 86 Vasile Preda - I.M.Stancu-Minasian - Anton Batatorescu, Optimality and duality in nonlinear programming involving semilocally preinvex and related functions.

Anno: **1995**

- ☐ n. 87 Elena Melis, Una nota storica sulla programmazione lineare: un problema di Kantorovich rivisto alla luce del problema degli zeri.
- ☐ n. 88 Vincenzo Bruno, Mobilità territoriale dell'Italia e di tre Regioni tipiche: Lombardia, Toscana, Sicilia.
- ☐ n. 89 Antonio Cortese, Bibliografia sulla presenza straniera in Italia
- ☐ n. 90 Riccardo Cambini, Funzioni scalari affini generalizzate.
- ☐ n. 91 Piero Manfredi - Fabio Tarini, Modelli epidemiologici: teoria e simulazione. (I)
- ☐ n. 92 Marco Bottai - Maria Caputo - Laura Lecchini, The "OLIVAR" survey.Methodology and quality.
- ☐ n. 93 Laura Lecchini - Donatella Marsiglia - Marco Bottai, Old people and social network.
- ☐ n. 94 Gilberto Ghilardi, Uno studio empirico sul confronto tra alcuni indici statistici della congiuntura.
- ☐ n. 95 Vincenzo Bruno, Il traffico nei porti italiani negli anni recenti.
- ☐ n. 96 Alberto Cambini - Anna Marchi - Laura Martein - S. Schaible, An analysis of the falk-palocsay algorithm.
- ☐ n. 97 Alberto Cambini - Laura Carosi, Sulla esistenza di elementi massimali.

Anno: **1996**

- ☐ n. 98 Riccardo Cambini - S. Komlòsi, Generalized concavity and generalized monotonicity concepts for vector valued.
- ☐ n. 99 Riccardo Cambini, Second order optimality conditions in the image space.
- ☐ n. 100 Vincenzo Bruno, La stagionalità delle correnti di navigazione marittima.
- ☐ n. 101 Eugene Maurice Cleur, A comparison of alternative discrete approximations of the Cox -I ngersoll - ross model.
- ☐ n. 102 Gilberto Ghilardi, Sul calcolo del rapporto di concentrazione del Gini.
- ☐ n. 103 Alberto Cambini - Laura Martein - Riccardo Cambini, A new approach to second order optimality conditions in vector optimization.
- ☐ n. 104 Fausto Gozzi, Alcune osservazioni sull'immunizzazione semideterministica.
- ☐ n. 105 Emilio Barucci - Fausto Gozzi, Innovation and capital accumulation in a vintage capital model an infinite dimensional control approach.
- ☐ n. 106 Alberto Cambini - Laura Martein - I.M.Stancu-Minasian., A survey of bicriteria fractional problems.
- ☐ n. 107 Luciano Fanti - Piero Manfredi, Viscosità dei salari, offerta di lavoro endogena e ciclo.
- ☐ n. 108 Piero Manfredi - Luciano Fanti, Ciclo di vita di nuovi prodotti: modellistica non lineare.
- ☐ n. 109 Piero Manfredi, Crescita con ciclo, gestazione dei piani di investimento ed effetti.
- ☐ n. 110 Luciano Fanti - Piero Manfredi, Un modello "classico" di ciclo con crescita ed offerta di lavoro endogena.
- ☐ n. 111 Anna Marchi, On the connecttedness of the efficient frontier : sets without local maxima.

☐ **n. 112** Riccardo Cambini, Generalized concavity for bicriteria functions.

☐ **n. 113** Vincenzo Bruno, Variazioni dinamiche (1971-1981-1991)dei fenomeni demografici dei comuni (urbani e rurali) della Lombardia, in relazione ad alcune caratteristiche di mobilità territoriale.

Anno: **1997**

☐ **n. 114** Piero Manfredi - Fabio Tarini - J.R. Williams - A. Carducci - B. Casini, Infectious diseases: epidemiology, mathematical models, and immunization policies.

☐ **n. 115** Eugene Maurice Cleur - Piero Manfredi, One dimensional SDE models, low order numerical methods and simulation based estimation: a comparison of alternative estimators.

☐ **n. 116** Luciano Fanti - Piero Manfredi, Point stability versus orbital stability (or instability): remarks on policy implications in classical growth cycle model.

☐ **n. 117** Piero Manfredi - Francesco Billari, transition into adulthood, marriage, and timing of life in a stable popolation framework.

☐ **n. 118** Laura Carosi, Una nota sul concetto di estremo superiore di insiemi ordinati da coni convessi.

☐ **n. 119** Laura Lecchini - Donatella Marsiglia, Reti sociali degli anziani: selezione e qualità delle relazioni.

☐ **n. 120** Piero Manfredi - Luciano Fanti, Gestation lags and efficiency wage mechanisms in a goodwin type growth model.

☐ **n. 121** G.Rivellini, La metodologia statistica multilevel come possibile strumento per lo studio delle interazioni tra il comportamento procreativo individuale e il contesto

☐ **n. 122** Laura Carosi, Una nota sugli insiemi C-limitati e L-limitati.

☐ **n. 123** Laura Carosi, Sull'estremo superiore di una funzione lineare fratta ristretta ad un insieme chiuso e illimitato.

☐ **n. 124** Piero Manfredi, A demographic framework for the evaluation of the impact of imported infectious diseases.

☐ **n. 125** Alessandro Valentini, Calo della fecondità ed immigrazione: scenari e considerazioni sul caso italiano.

☐ **n. 126** Alberto Cambini - Laura Martein, Second order optimality conditions.

Anno: **1998**

☐ **n. 127** Piero Manfredi and Alessandro Valentini, Populations with below replacementfertility: theoretical considerations and scenarioes from the italian laboratory.

☐ **n. 128** Alberto Cambini - Laura Martein - E. Moretti, Programmazione frazionaria e problemi bicriteria.

☐ **n. 129** Emilio Barucci - Fausto Gozzi - Andrej Swiech, Incentive compatibility constraints and dynamic programming in continuous time.

Anno: **1999**

☐ **n. 130** Alessandro Valentini, Impatto delle immigrazioni sulla popolazione italiana: confronto tra scenari alternativi.

☐ **n. 131** K. Iglicka - Odo Barsotti - Laura Lecchini, Recent developement of migrations from Poland to Europe with a special emphsis on Italy K.Iglika - Le Migrazioni est-ovest: le unioni miste in Italia

☐ **n. 132** Alessandro Valentini, Proiezioni demografiche multiregionali a due sessi, con immigrazioni internazionali e vincoli di consistenza.

☐ **n. 133** Fabio Antonelli - Emilio Barucci - Maria Elvira Mancino, Backward-forward stochastic differential utility: existence, consumption and equilibrium analysis.

☐ **n. 134** Emilio Barucci - Maria Elvira Mancino, Asset pricing with endogenous aspirations.

☐ **n. 135** Eugene Maurice Cleur, Estimating a class of diffusion models: an evaluation of the effects of sampled discrete observations.

☐ **n. 136** Luciano Fanti - Piero Manfredi, Labour supply, time delays, and demoeconomic oscillations in a solow-typegrowth model.

☐ **n. 137** Emilio Barucci - Sergio Polidoro - Vincenzo Vespri, Some results on partial differential equations and Asian options.

☐ **n. 138** Emilio Barucci - Maria Elvira Mancino, Hedging european contingent claims in a Markovian incomplete market.

☐ **n. 139** Alessandro Valentini, L'applicazione del modello multiregionale-multistato alla popolazione in Italia mediante l'utilizzo del Lipro: procedura di adattamento dei dati e particolarità tecniche del programma.

☐ **n. 140** I.M.Stancu-Minasian, optimality conditions and duality in fractional programming-involving semilocally preinvex and related functions.

☐ **n. 141** Alessandro Valentini, Proiezioni demografiche con algoritmi di consistenza per la popolazione in Italia nel periodo 1997-2142: presentazione dei risultati e confronto con metodologie di stima alternative.

☐ **n. 142** Laura Carosi, Competitive equilibria with money and restricted partecipation.

☐ **n. 143** Laura Carosi, Monetary policy and Pareto improvability in a financial economy with restricted partecipation

☐ **n. 144** Bruno Cheli, Misurare il benessere e lo sviluppo dai paradossi del Pil a misure di benessere economico sostenibile, con uno sguardo allo sviluppo umano

☐ **n. 145** Bruno Cheli - Laura Lecchini - Lucio Masserini, The old people's perception of well-being: the role of material and non material resources

☐ **n. 146** Eugene Maurice Cleur, Maximum likelihood estimation of one-dimensional stochastic differential equation models from discrete data: some computational resultas

☐ **n. 147** Alessandro Valentini - Francesco Billari - Piero Manfredi, Utilizzi empirici di modelli multistato continui con durate multiple

☐ **n. 148** Francesco Billari - Piero Manfredi - Alberto Bonaguidi - Alessandro Valentini, Transition into adulthoold: its macro-demographic consequences in a multistatew stable population framework

☐ **n. 149** Francesco Billari - Piero Manfredi - Alessandro Valentini, Becoming Adult and its Macro-Demographic Impact: Multistate Stable Population Theory and an Application to Italy

☐ **n. 150** Alessandro Valentini, Le previsioni demografiche in presenza di immigrazioni: confronto tra modelli alternativi e loro utilizzo empirico ai fini della valutazione dell'equilibrio nel sistema pensionistico

☐ **n. 151** Emilio Barucci - Roberto Monte, Diffusion processes for asset prices under bounded rationality

☐ **n. 152** Emilio Barucci - P. Cianchi - L. Landi - A. Lombardi, Reti neurali e analisi delle serie storiche: un modello per la previsione del BTP future

☐ **n. 153** Alberto Cambini - Laura Carosi - Laura Martein, On the supremum in fractional programming

☐ **n. 154** Riccardo Cambini - Laura Martein, First and second order characterizations of a class of pseudoconcave vector functions

☐ **n. 155** Piero Manfredi and Luciano Fanti, Embedding population dynamics in macro-economic models. The case of the goodwin's growth cycle

☐ **n. 156** Laura Lecchini e Odo Barsotti, Migrazioni dei preti dalla Polonia in Italia

☐ **n. 157** Vincenzo Bruno, Analisi dei prezzi, in Italia dal 1975 in poi

☐ **n. 158** Vincenzo Bruno, Analisi del commercio al minuto in Italia

☐ **n. 159** Vincenzo Bruno, Aspetti ciclici della liquidità bancaria, dal 1971 in poi

☐ **n. 160** Anna Marchi, A separation theorem in alternative theorems and vector optimization

Anno: **2000**

- ☐ **n. 161** Piero Manfredi and Luciano Fanti, Labour suppley, population dynamicics and persistent oscillations in a Goodwin-type growth cycle model
- ☐ **n. 162** Luciano Fanti and Piero Manfredi, Neo-classical labour market dynamics and chaos (and the Phillips curve revisited)
- ☐ **n. 163** Piero Manfredi - and Luciano Fanti, Detection of Hopf bifurcations in continuos-time macro- economic models, with an application to reducible delay-systems.
- ☐ **n. 164** Fabio Antonelli - Emilio Barucci, The Dynamics of pareto allocations with stochastic differential utility
- ☐ **n. 165** Eugene M. Cleur, Computing maximum likelihood estimates of a class of One-Dimensional stochastic differential equation models from discrete Date*
- ☐ **n. 166** Eugene M. Cleur, Estimatng the drift parameter in diffusion processes more efficiently at discrete times:a role of indirect estimation
- ☐ **n. 167** Emilio Barucci - Vincenzo Valori, Forecasting the forecasts of others e la Politica di Inflation targeting
- ☐ **n. 168** A.Cambini - L. Martein, First and second order optimality conditions in vector optimization
- ☐ **n. 169** A. Marchi, Theorems of the Alternative by way of Separation Theorems
- ☐ **n. 170** Emilio Barucci - Maria Elvira Mancino, Asset Pricing and Diversification with Partally Exchangeable random Variables
- ☐ **n. 171** Piero Manfredi - Luciano Fanti, Long Term Effects of the Efficiency Wage Hypothesis in Goodwin-Type Economies.
- ☐ **n. 172** Piero Manfredi - Luciano Fanti, Long Term Effects of the Efficiency wage Hypothesis in Goodwin-type Economies: a reply.
- ☐ **n. 173** Luciano Fanti, Innovazione Finanziaria e Domanda di Moneta in un Modello dinamico IS-LM con Accumulazione.
- ☐ **n. 174** P.Manfredi, A.Bonaccorsi, A.Secchi, Social Heterogeneities in Classical New Product Diffusion Models. I: "External" and "Internal" Models.
- ☐ **n. 175** Piero Manfredi - Ernesto Salinelli, Modelli per formazione di coppie e modelli di Dinamica familiare.
- ☐ **n. 176** P.Manfredi, E. Salinelli, A.Melegaro, A.Secchi, Long term Interference Between Demography and Epidemiology: the case of tuberculosis
- ☐ **n. 177** Piero Manfredi - Ernesto Salinelli, Toward the Development of an Age Structure Teory for Family Dynamics I: General Frame.
- ☐ **n. 178** Piero Manfredi - Luciano Fanti, Population heterogeneities, nonlinear oscillations and chaos in some Goodwin-type demo-economic models Paper to be presented at the: Second workshop on "nonlinear demography" Max Planck Institute for demographic Research Rostock, Germany, May 31-June 2, 2
- ☐ **n. 179** E. Barucci - M.E. Mancini - Roberto Renò, Volatility Estimation via Fourier Analysis
- ☐ **n. 180** Riccardo Cambini, Minimum Principle Type Optimality Conditions
- ☐ **n. 181** E. Barucci, M. Giuli, R. Monte, Asset Prices under Bounded Rationality and Noise Trading
- ☐ **n. 182** A. Cambini, D.T.Luc, L.Martein, Order Preserving Transformations and application.
- ☐ **n. 183** Vincenzo Bruno, Variazioni dinamiche (1971-1981-1991) dei fenomeni demografici dei comuni urbani e rurali della Sicilia, in relazione ad alcune caratteristiche di mobilità territoriale.
- ☐ **n. 184** F.Antonelli, E.Barucci, M.E.Mancino, Asset Pricing with a Backward-Forward Stochastic Differential Utility
- ☐ **n. 185** Riccardo Cambini - Laura Carosi, Coercivity Concepts and Recession Functions in Constrained Problems
- ☐ **n. 186** John R. Williams, Piero Manfredi, The pre-vaccination dynamics of measles in Italy: estimating levels of under-reporting of measles cases
- ☐ **n. 187** Piero Manfredi, John R. Williams, To what extent can inter-regional migration perturbe local endemic patterns? Estimating numbers of measles cases in the Italian regions
- ☐ **n. 188** Laura Carosi, Johannes Jahn, Laura Martein, On The Connections between Semidefinite Optimization and Vector Optimization
- ☐ **n. 189** Alberto Cambini, Jean-Pierre Crouzeix, Laura Martein, On the Pseudoconvexity of a Quadratic Fractional Function
- ☐ **n. 190** Riccardo Cambini - Claudio Sodini, A finite Algorithm for a Particular d.c. Quadratic Programming Problem.
- ☐ **n. 191** Riccardo Cambini - Laura Carosi, Pseudoconvexity of a class of Quadratic Fractional Functions.
- ☐ **n. 192** Laura Carosi, A note on endogenous restricted partecipation on financial markets: an existence result.
- ☐ **n. 193** Emilio Barucci - Roberto Monte - Roberto Renò, Asset Price Anomalies under Bounded Rationality.
- ☐ **n. 194** Emilio Barucci - Roberto Renò, A Note on volatility estimate-forecast with GARCH models.
- ☐ **n. 195** Bruno Cheli, Sulla misura del benessere economico: i paradossi del PIL e le possibili correzioni in chiave etica e sostenibile, con uno spunto per l'analisi della povertà
- ☐ **n. 196** M.Bottai, M.Bottai, N. Salvati, M.Toigo, Le proiezioni demografiche con il programma Nostradamus. (Applicazione all'area pisana)
- ☐ **n. 197** A. Lemmi - B. Cheli - B. Mazzolli, La misura della povertà multidimensionale: aspetti metodologici e analisi della realtà italiana alla metà degli anni '90
- ☐ **n. 198** C.R. Bector - Riccardo Cambini, Generalized B-invex vector valued functions
- ☐ **n. 199** Luciano Fanti - Piero Manfredi, The workers' resistance to wage cuts is not necessarily detrimental for the economy: the case of a Goodwin's growth model with endogenous population.
- ☐ **n. 200** Emilio Barucci - Roberto Renò, On Measuring volatility of diffusion processes with high frequency data
- ☐ **n. 201** Piero Manfredi - Luciano Fanti, Demographic transition and balanced growth

Anno: **2001**

- ☐ **n. 202** E.Barucci - M. E. Mancini - E. Vannucci, Asset Pricing, Diversification and Risk Ordering with Partially Exchangeable random Variables
- ☐ **n. 203** E. Barucci - R. Renò - E. Vannucci, Executive Stock Options Evaluation.
- ☐ **n. 204** Odo Barsotti - Moreno Toigo, Dimensioni delle rimesse e variabili esplicative: un'indagine sulla collettività marocchina immigrata nella Toscana Occidentale
- ☐ **n. 205** Vincenzo Bruno, I Consumi voluttuari, nell'ultimo trentennio, in Italia
- ☐ **n. 206** Michele Longo, The monopolist choice of innovation adoption: A regular-singular stochastic control problem
- ☐ **n. 207** Michele Longo, The competitive choice of innovation adoption: A finite-fuel singular stochastic control problem.
- ☐ **n. 208** Riccardo Cambini - Laura Carosi, On the pseudoaffinity of a class of quadratic fractional functions
- ☐ **n. 209** Riccardo Cambini - Claudio Sodini, A Finite Algorithm for a Class of Non Linear Multiplicative Programs.
- ☐ **n. 210** Alberto Cambini - Dinh The Luc - Laura Martein, A method for calculating subdifferential Convex vector functions
- ☐ **n. 211** Alberto Cambini - Laura Martein, Pseudolinearity in scalar and vector optimization.
- ☐ **n. 212** Riccardo Cambini, Necessary Optimality Conditions in Vector Optimization.
- ☐ **n. 213** Riccardo Cambini - Laura Carosi, On generalized convexity of quadratic fractional functions.
- ☐ **n. 214** Riccardo Cambini - Claudio Sodini, A note on a particular quadratic programming problem.
- ☐ **n. 215** Michele Longo - Vincenzo Valori, Existence end stability of equilibria in OLG models under adaptive expectations.

- ☐ **n. 216** Luciano Fanti - Piero Manfredi, Population, unemployment and economic growth cycles: a further explanatory perspective
- ☐ **n. 217** J.R.Williams,P.Manfredi,S.Salmaso,M.Ciofi, Heterogeneity in regional notification patterns and its impact on aggregate national case notification data: the example of measles in Italy.
- ☐ **n. 218** Anna Marchi, On the connectedness of the efficient frontier: sets without local efficient maxima
- ☐ **n. 219** Laura Lecchini - Odo Barsotti, Les disparités territoriales au Maroc au travers d'une optique de genre.

Anno: **2002**

- ☐ **n. 220** Gilberto Ghilardi - Nicola Orsini, Sull'uso dei modelli statistici lineari nella valutazione dei sistemi formativi.
- ☐ **n. 221** Andrea Mercatanti, Un'analisi descrittiva dei laureati dell'Università di Pisa
- ☐ **n. 222** E. Barucci - C. Impenna - R. Renò, The Italian Overnight Market: microstructure effects, the martingale hypothesis and the payment system.
- ☐ **n. 223** E. Barucci, P.Malliavin, M.E.Mancino, R.Renò, A.Thalmaier, The Price-volatility feedback rate: an implementable mathematical indicator of market stability.
- ☐ **n. 224** Andrea Mercatanti, Missing at random in randomized experiments with imperfect compliance
- ☐ **n. 225** Andrea Mercatanti, Effetto dell'uso di carte Bancomat e carte di Credito sulla liquidità familiare: una valutazione empirica
- ☐ **n. 226** Piero Manfredi - John R. Williams, Population decline and population waves: their impact upon epidemic patterns and morbidity rates for childhood infectious diseases. Measles in Italy as an example.
- ☐ **n. 227** Piero Manfredi - Marta Ciofi degli Atti, La geografia pre-vaccinale del morbillo in Italia. I. Comportamenti di contatto e sforzi necessari all'eliminazione: predizioni dal modello base delle malattie prevenibili da vaccino.
- ☐ **n. 228** I.M.Stancu-Minasian, Optimality Conditions and Duality in Fractional Programming Involving Semilocally Preinvex and Related
- ☐ **n. 229** Nicola Salvati, Un software applicativo per un'analisi di dati sui marchi genetici (Genetic Markers)
- ☐ **n. 230** Piero Manfredi, J. R. Williams, E. M. Cleur, S. Salmaso, M. Ciofi, The pre-vaccination regional landscape of measles in Italy: contact patterns and related amount of needed eradication efforts (and the "EURO" conjecture)
- ☐ **n. 231** Andrea Mercatanti, I tempi di laurea presso l'Università di Pisa: un'applicazione dei modelli di durata in tempo discreto
- ☐ **n. 232** Andrea Mercatanti, The weak version of the exclusion restriction in causal effects estimation: a simulation study
- ☐ **n. 233** Riccardo Cambini and Laura Carosi, Duality in multiobjective optimization problems with set constraints
- ☐ **n. 234** Riccardo Cambini and Claudio Sodini, Decomposition methods for nonconvex quadratic programs
- ☐ **n. 235** R.Cambini and L. Carosi and S.Schaible, Duality in fractional optimization problems with set constraints
- ☐ **n. 236** Anna Marchi, On the mix-efficient points

Anno: **2003**

- ☐ **n. 237** Emanuele Vannucci, The valuation of unit linked policies with minimal return guarantees under symmetric and asymmetric information hypoteses
- ☐ **n. 238** John R Williams - Piero Manfredi, Ageing populations and childhood infections: the potential impact on epidemic patterns and morbidity
- ☐ **n. 239** Bruno Cheli, Errata Corrige del Manuale delle Impronte Ecologiche (2002) ed alcuni utili chiarimenti
- ☐ **n. 240** Alessandra Petrucci-Nicola Salvati-Monica Pratesi, Stimatore Combinato r Correlazione Spaziale nella Stima per Piccole Aree
- ☐ **n. 241** Riccardo Cambini - Laura Carosi, Mixed Type Duality for Multiobjective Optimization Problems with set constraints
- ☐ **n. 242** O.Barsotti, L.Lecchini, F.Benassi, Foreigners from central and eastern European countries in Italy: current and future perspectives of eu enlargement
- ☐ **n. 243** A. Cambini - L. Martein - S. Schaible, Pseudoconvexity under the Charnes-Cooper transformation
- ☐ **n. 244** Eugene M. Cleur, Piero Manfredi, and John R. William, The pre-and post-Vaccination regional dynamics of measles in Italy: Insights from time series analysis

Anno: **2004**

- ☐ **n. 245** Emilio Barucci - Jury Falini, Determinants of Corporate Governance in Italy: Path dependence or convergence?
- ☐ **n. 246** R. Cambini - A. Marchi, A note on the connectedness of the efficient frontier
- ☐ **n. 247** Laura Carosi - Laura Martein, On the pseudoconvexity and pseudolinearity of some classes of fractional functions
- ☐ **n. 248** E. Barucci - R. Monte - B. Trivellato, Bayesian nash equilibrium for insider trading in continuous time
- ☐ **n. 249** Eugene M. Cleur, A Time Series Analysis of the Inter-Epidemic Period for Measles in Italy
- ☐ **n. 250** Andrea Mercatanti, Causal inference methods without exclusion restrictions: an economic application.
- ☐ **n. 251** Eugene M. Cleur, Non-Linearities in Monthly Measles data for Italy
- ☐ **n. 252** Eugene M. Cleur, A Treshold Model for Prevaccination Measles Data: Some Empirical Results for England and Italy
- ☐ **n. 253** Andrea Mercatanti, La gestione dei dati mancanti nei modelli di inferenza causale: il caso degli esperimenti naturali.
- ☐ **n. 254** Andrea Mercatanti, Rilevanza delle analisi di misture di distribuzioni nelle valutazioni di efficacia
- ☐ **n. 255** Andrea Mercatanti, Local estimation of mixtures in instrumental variables models
- ☐ **n. 256** Monica Pratesi - Nicola Salvati, Spatial EBLUP in agricultural surveys: an application based on italian census data.
- ☐ **n. 257** Emanuele Vannucci, A model analyzing the effects of information asymmetries of the traders
- ☐ **n. 258** Monica Pratesi-Emilia Rocco, Two-Step centre sampling for estimating elusive population size
- ☐ **n. 259** A.Lemmi, N.Pannuzi, P.Valentini, B.Cheli, G.Berti, Estimating Multidimensional Poverty: A Comparison of Three Diffused Methods°

Anno: **2005**

- ☐ **n. 260** Nicola Salvati, Small Area estimation: the EBLUP estimator using the CAR model
- ☐ **n. 261** Monica Pratesi-Nicola Salvati, Small Area Estimation: the EBLUP estimator with autoregressive random area effects
- ☐ **n. 262** Riccardo Cambini-Claudio Sodini, A solution algorithm for a class of box constrained quadratic programming problems
- ☐ **n. 263** Andrea Mercatanti, A constrained likelihood maximization for relaxing the exclusion restriction in causal inference.
- ☐ **n. 264** Marco Bottai - Annalisa Lazzini - Nicola Salvati, Le proiezioni demografiche. Pisa 2003/2032
- ☐ **n. 265** Andrea Mercatanti, An exercise in estimating causal effects for non-compliers: the return to schooling in Germany and Austria
- ☐ **n. 266** Nicola Salvati, M-quantile Geographically Weighted Regression for Nonparametric Small Area Estimation
- ☐ **n. 267** Ester Rizzi, Alessandro Rosina, L'influsso della Luna sul comportamento sessuale
- ☐ **n. 268** Silvia Venturi, Linda Porciani, Moreno Toigo, Federico Benassi, Il migrate nello spazio sociale transnazionale: tra integrazione nel Paese di