

Optimal column density measurements from multiband near-infrared observations[★]

M. Lombardi^{1,2}

¹ European Southern Observatory, Karl-Schwarzschild-Straße 2, 85748 Garching bei München, Germany
e-mail: mlombard@eso.org

² University of Milan, Department of Physics, via Celoria 16, 20133 Milan, Italy

Received 27 October 2004 / Accepted 22 March 2005

Abstract. We consider from a general point of view the problem of determining the extinction in dense molecular clouds. We use a rigorous statistical approach to characterize the properties of the most widely used optical and infrared techniques, namely the star count and the color excess methods. We propose a new maximum-likelihood method that takes advantage of both star counts and star colors to provide an optimal estimate of the extinction. Detailed numerical simulations show that our method performs optimally under a wide range of conditions and, in particular, is significantly superior to the standard techniques for clouds with high column-densities and affected by contamination by foreground stars.

Key words. dust, extinction – methods: statistical – ISM: clouds – infrared: ISM – stars: formation

1. Introduction

A detailed comprehension of the structure and physical properties of dark molecular clouds is a critical step to understand fundamental processes such as star and planet formation. Still, after several decades of investigations, very little is known about dark clouds, about their relationship to the diffuse interstellar medium, and about the composition and evolution of the dust grains present in the clouds.

Molecular clouds are composed of approximately 99% molecular hydrogen and helium, but due to the absence of a dipole moment these molecules are virtually undetectable at the low temperature (~ 10 K) that characterize these objects. As a result, astronomers have been using different tracers to study the density distribution of molecular clouds. Historically, the first technique was based on a statistical analysis of the angular density of stars observed through a cloud (Wolf 1923; Bok 1937). This method, known as star counts, uses the dust (which is responsible for the extinction of light) as a tracer of H_2 and He.

More recently, radio observations of carbon monoxide (CO) and its isotopes have been used to study the spatial distribution and the physical conditions of molecular clouds, under the assumption that CO is a reliable tracer of the matter in the cloud, i.e. that the ratio $N(\text{CO})/(N(\text{H}^2) + N(\text{He}))$ is approximately constant. Radio spectroscopy techniques have provided extremely effective in studying the most dense regions (above 10^{22} protons cm^{-2}) and, more importantly, give

dynamical information on the cloud structure. However, with the advent of various dust detectors in the near-infrared (NIR), far infrared, millimeter, and sub-millimeter bands, it has become clear that several poorly constrained processes (e.g., deviations from thermodynamic equilibrium, opacity variations, chemical evolution, and more importantly depletion of molecules) can significantly affect the results of analyses based on radio observations (e.g. Lada et al. 1994; Alves et al. 1999; Harjunpää et al. 2004).

The reddening of background stars offers a natural method to study the distribution of dust in molecular clouds, and thus the hydrogen column density. This technique can be better applied to the infrared bands, which compared to optical bands are less affected by extinction and are less sensitive to the physical properties of the dust grains (Mathis 1990). Before the advent of large format array cameras, the lack of instrumental sensitivity clogged infrared observations to small, dense clouds (e.g. Jones et al. 1980; Frerking et al. 1982; Jones et al. 1984; Casali 1986). More recently, infrared arrays have made it possible to measure thousands of stars and to extend the original technique to entire molecular cloud complexes (Lada et al. 1994, 1999; Alves et al. 2001; Lombardi & Alves 2001). Such measurements are free of the complications that plague molecular-line data and permit a detailed analysis of the cloud density distribution.

Although the success of the NIR color excess method is evident, there is still a need for a deeper understanding of its limitations, statistical biases, and uncertainties. The aim of this paper is twofold: (i) to study in detail the statistical properties of the star count and color excess methods and (ii) to describe a

[★] Appendix A is only available in electronic form at <http://www.edpsciences.org>

new, optimal method based on a maximum-likelihood analysis. The method is described here for NIR observations, but could equally well be applied to other infrared bands for which the extinction law and the intrinsic color distribution are known with good accuracy. The paper is organized as follows. In Sect. 2 we introduce our formalism and consider from a statistical point of view the extinction of stars by a foreground cloud. The two main techniques used to obtain extinction measurements, namely the star count and the NIR color excess methods, are discussed in Sect. 3. In Sect. 4 we describe the new maximum-likelihood method and show by means of numerical simulations that it performs excellently even in the presence of a significant contamination by foreground stars. Our conclusions are briefly reported in Sect. 5. Finally, in Appendix A we consider the statistical properties of the median and of related estimators used often in infrared studies of dark clouds to remove the effects of foreground stars.

2. Basic relations

Suppose that N stars with magnitudes (on a given band) $\{\hat{m}_n\}$ are observed in a given region of the sky (hereafter we will use the hat $\hat{}$ to denote measured quantities). These observed magnitudes depend on the original, unreddened, absolute star magnitudes $\{M_n\}$, on the star distances $\{D_n\}$, on the individual extinctions $\{A_n\}$, and on the photometric errors $\{\epsilon_n\}$. In particular, we have

$$\hat{m}_n = M_n + 5 \log_{10} D_n - 5 + A_n + \epsilon_n = m_n + \epsilon_n, \quad (1)$$

where the distances $\{D_n\}$ are taken to be expressed in parsecs. The quantity m_n that appears in the r.h.s. of this equation is the reddened magnitude of the n th star free from measurement errors (i.e., the magnitude that would be observed in the limit of an extremely long exposure time).

2.1. Single-band probability distributions

In order to consider the most general situation in a proper statistical way, we introduce several probability distributions:

$\rho(M, D)$: the probability distribution (density) of stars with absolute magnitude M at distance D ;

$p_A(A|D)$: the probability distribution of extinction for objects located at distance D , i.e. the probability of an extinction of A magnitudes on a star given that its distance is D ;

$p_\epsilon(\hat{m}|m)$: the probability distribution of photometric measurement errors, i.e. the probability of measuring a magnitude \hat{m} for a star given that its true reddened magnitude is m .

These probability distributions are now defined in the most general way. Later, however, we will consider special forms of them that allow us to simplify the relevant equations. Note also that we take the extinction A at distance D as a random variable. Hence, we have introduced above the probability distribution $p_A(A|D)$ rather than a (deterministic) function $A(D)$. This general approach can be used to describe molecular clouds with ‘‘patchy’’ column densities, which seems to be quite common (see, e.g., Lada et al. 1999; Cambr esly 1999). Finally, we

use a general form for the photometric error that includes the common case where the error depends on the magnitude of the star. Moreover, we also consider through $p_\epsilon(\hat{m}|m)$ the case of *undetected* objects, i.e. the case where a star of true magnitude m close to the detection limit produces no detectable flux; in this case we write $\hat{m} = \text{null}$. Hence, the quantity

$$c(m) = 1 - p_\epsilon(\text{null}|m) \quad (2)$$

represents the *completeness* of our observations for stars with true magnitude m .

We assume that both $p_A(A|D)$ and $p_\epsilon(\hat{m}|m)$ are normalized to unity, i.e.

$$\int_0^\infty p_A(A|D) dA = 1 \quad \forall D, \quad (3)$$

$$p_\epsilon(\text{null}|m) + \int_{-\infty}^\infty p_\epsilon(\hat{m}|m) d\hat{m} = 1 \quad \forall m. \quad (4)$$

Moreover, we take $\rho(M, D)$ to be normalized so that the quantity

$$dN = D^2 \rho(M, D) dD d\Omega dM \quad (5)$$

represents the expected number of stars located at a distance in the range $[D, D + dD]$ in a patch of the sky of solid angle $d\Omega$, and with absolute magnitudes between M and $M + dM$.

Equation (1) can be used to obtain the probability distribution of reddened magnitudes $p_m(m)$. Using the previous definitions we find

$$p_m(m) = \int_0^\infty dD D^2 \int_0^\infty dA p_A(A|D) \times \rho(m - 5 \log_{10} D + 5 - A, D). \quad (6)$$

Finally, $p_m(m)$ can be converted into the distribution of observed magnitudes $p_{\hat{m}}(\hat{m})$ by convolving it with the measurement error probability distribution $p_\epsilon(\hat{m}|m)$:

$$p_{\hat{m}}(\hat{m}) = \int_{-\infty}^\infty p_\epsilon(\hat{m}|m) p_m(m) dm. \quad (7)$$

Because of the way $p_\epsilon(\hat{m}|m)$ is defined, Eq. (7) includes also the case $\hat{m} = \text{null}$. Note also that $p_{\hat{m}}(\hat{m})$ is normalized so that $p_{\hat{m}}(\hat{m}) d\hat{m}$ represents the expected angular density of stars with observed magnitudes on the range $[\hat{m}, \hat{m} + d\hat{m}]$.

Equations (6) and (7) are basic equations of this paper. However, in the form they are written, they are by far too general to be useful in practical cases.

2.2. Simplifications

In order to make use of the general relations written above we introduce a number of useful simplifications.

First, we note that so far we have considered observations carried out on a *single* band. In order to treat multi-band data, we introduce a *reddening law* as follows: for any band λ , the extinction A_λ is given by

$$A_\lambda = k_\lambda A_V, \quad (8)$$

where k_λ is a constant and A_V is the extinction in the V band. In other words, Eq. (8) states that the ratio $A_\lambda/A_{\lambda'}$ between the

extinction in two bands λ and λ' is a constant for a given cloud. In reality, this ratio is known to vary with the environment for short wavelengths, but is almost universal in the NIR (and, according to a recent work of Indebetouw et al. 2004, for wavelengths in the range 3–10 μm as well). In the rest of this paper we will follow the usual notation and express all extinctions in terms of A_V .

A second important assumption that we will use regards the functional form of $\rho(M, D)$. First, we will assume that this function can be separated in M and D , i.e. can be written as the product of two functions, one involving M only, and one involving D only: $\rho(M, D) = p_M(M)\rho(D)$. In practice, this assumption is equivalent to saying that we observe a single “population” of stars at all distances. Since observations show that star number counts are well approximated by an exponential law of the form $p_m(m) \propto 10^{\alpha m}$ (where the exponent $\alpha \simeq 0.34$ is approximately independent of the band considered in the NIR), we also assume that $p_M(M)$ follows a similar law. We then write

$$\rho(M, D) = \nu_\lambda 10^{\alpha M} \rho(D), \quad (9)$$

where ν_λ is a normalization factor. In the following we will assume that $\rho(D)$, similarly to α , is independent of the band considered, while ν_λ is not. Note also that, because of the exponential form of this equation in M , a change of the normalization factor ν_λ is in practice equivalent to a change of the zero-point used for the magnitude M .

Using Eqs. (9) and (8) we can slightly simplify Eq. (6) by performing the integration over D :

$$\begin{aligned} p_m(m) &= \int_0^\infty dD D^{2-5\alpha} \rho(D) \int_0^\infty dA_V p_{A_V}(A_V|D) \\ &\quad \times \nu_\lambda 10^{5\alpha} 10^{\alpha m} 10^{-\alpha k_\lambda A_V} \\ &\equiv a \nu_\lambda 10^{\alpha m} \int_0^\infty dA_V p_{A_V}(A_V) 10^{-\alpha k_\lambda A_V}, \end{aligned} \quad (10)$$

where we have defined a distance-weighted probability distribution for A_V as

$$p_{A_V}(A_V) = \frac{10^{5\alpha}}{a} \int_0^\infty dD D^{2-5\alpha} \rho(D) p_{A_V}(A_V|D), \quad (11)$$

and where a is a numerical factor introduced in Eq. (10) to ensure that $p_{A_V}(A_V)$ is normalized to unity:

$$a = 10^{5\alpha} \int_0^\infty dD D^{2-5\alpha} \rho(D). \quad (12)$$

Note that $p_{A_V}(A_V)$ has a simple interpretation, being the probability distribution for stars to have undergone a given extinction regardless of their distance. For example, if all stars were subject to the same extinction A_V , we would have $p_{A_V}(A'_V) = \delta(A'_V - A_V)$. Note also that, since $\alpha \simeq 0.34$ can be taken to be approximately independent of the band considered, $p_{A_V}(A_V)$ is also band-independent.

Equation (10) shows an interesting property of extinction studies: the probability distribution $p_m(m)$ depends on $p_{A_V}(A_V|D)$ only through $p_{A_V}(A_V)$. This point is particularly important since it shows an intrinsic limitation of extinction measurements. Indeed, all observables are derived from $p_{\hat{m}}(\hat{m})$,

the distribution of observed magnitudes, and this function depends only on $p_m(m)$ (other than observational “limiting factors” such as $p_e(\hat{m}|m)$ or $\hat{c}(\hat{m})$). Since, as noted above, $p_m(m)$ does not depend directly on $p_{A_V}(A_V|D)$, any estimator based on observed magnitudes only¹ will not provide any information on $p_{A_V}(A_V|D)$ directly but (in the best case scenario) only on $p_{A_V}(A_V)$. Since $p_{A_V}(A_V)$ is a sort of convolution of $p_{A_V}(A_V|D)$, in general it is not possible to have a complete knowledge on $p_{A_V}(A_V|D)$, not even if we know the star density distribution $\rho(D)$; this limitation, among other things, prevents us from gathering information on the three-dimensional structure of a molecular cloud.

A case where, instead, the function $p_{A_V}(A_V|D)$ can be recovered is a *deterministic model* for the extinction, i.e. a cloud complex whose extinction A_V depends directly on the distance D (and, thus, is not a random variable). In this case we have

$$p_{A_V}(A_V|D) = \delta(A_V - A_V(D)). \quad (13)$$

The extinction $A_V(D)$ has a simple relationship with the integrated gas column density ρ_{gas} :

$$A_V(D) = \frac{1}{\beta} \int_0^D dD' \rho_{\text{gas}}(D'), \quad (14)$$

where $\beta \simeq 2 \times 10^{21} \text{ cm}^{-2} \text{ mag}^{-1}$ is the ratio between gas density and V -band extinction (Lilley 1955; Bohlin et al. 1978). Suppose now that we carry out observations on the cloud and measure $p_{A_V}(A_V)$. In order to show that we can recover the structure of the cloud, let us call $D(A_V)$ the inverse of the function $A_V(D)$ defined in Eq. (14), i.e. the distance at which the integrated extinction is A_V . Then we have

$$\begin{aligned} p_{A_V}(A_V) &= \frac{10^{5\alpha}}{a} [D(A_V)]^{2-5\alpha} \rho(D(A_V)) D'(A_V) \\ &= \frac{10^{5\alpha}}{a} \frac{\rho(D(A_V))}{3-5\alpha} \frac{d}{dA_V} \left[[D(A_V)]^{3-5\alpha} \right]. \end{aligned} \quad (15)$$

In other words, we have obtained a differential equation in $[D(A_V)]^{3-5\alpha}$. If $\rho(D)$ is known, using the boundary condition $D(0) = 0$ we can in principle solve this differential equation and obtain $D(A_V)$ which, in turn, can be inverted into $A_V(D)$ and $\rho_{\text{gas}}(D)$. It is superfluous to stress that this process in reality is far from being trivial.

Finally, we consider a very simple deterministic model for $p_{A_V}(A_V|D)$ that describes a thin cloud at a distance D_0 with uniform column density A_V :

$$p_{A_V}(A'_V|D) = \begin{cases} \delta(A'_V) & \text{if } D < D_0, \\ \delta(A'_V - A_V) & \text{otherwise.} \end{cases} \quad (16)$$

The related distribution $p_{A_V}(A_V)$ takes then the simple form

$$p_{A_V}(A'_V) = f \delta(A'_V) + (1-f) \delta(A'_V - A_V), \quad (17)$$

where f is a real number in the range $[0, 1]$ given by

$$f = \left[\int_0^{D_0} dD D^{2-5\alpha} \rho(D) \right] \left/ \left[\int_0^\infty dD D^{2-5\alpha} \rho(D) \right] \right. . \quad (18)$$

¹ Through this paper we assume that the distance of individual stars is not an observable; if, instead, the distance of each star can be estimated, then in principle one can also directly measure $p_{A_V}(A_V|D)$.

Hence, f represents the fraction of foreground stars (i.e. stars at distance $D < D_0$) present in any apparent magnitude bin in regions with negligible extinction. In the following, we will refer to the simple case described in Eqs. (16) and (17) as “thin cloud approximation” (the term is borrowed from gravitational lensing theory).

2.3. The completeness function

Above, in Eq. (2), we introduced the completeness function $c(m)$, which gives the probability to *detect* a star of magnitude m . Typically, $c(m)$ is close to unity for bright stars, and vanishes for very faint stars; note however that $c(m)$ might be smaller than unity at low m in crowded fields or close to bright objects. The completeness function $c(m)$ enters naturally in the definition of the error probability distribution $p_\epsilon(\hat{m}|m)$, which can be written as

$$p_\epsilon(\hat{m}|m) = \begin{cases} c(m)p_\epsilon(\hat{m}|m) & \text{if } \hat{m} \neq \text{null} , \\ 1 - c(m) & \text{if } \hat{m} = \text{null} . \end{cases} \quad (19)$$

This representation reflects the measurement process: for any star, there is first a random process that “decides” whether the object is detected or not (with probability fixed by $c(m)$); then, if the star is detected, there is a second random process that generates the observed magnitude \hat{m} according to $p_\epsilon(\hat{m}|m)$. Note also that this latter probability distribution is normalized (cf. Eq. (4)).

In order to carry out some simplifications (see below Sect. 4.2), we also consider a different representation of the completeness function in which the order of the two random processes described above is swapped: this leads to a completeness function $\hat{c}(\hat{m})$ defined in terms of the *observed* magnitude. In other words, we first generate for *every* star the “observed” magnitude \hat{m} according to a probability distribution $\hat{p}_\epsilon(\hat{m}|m)$, and then decide whether the star is really observed using a second random process controlled by the completeness function $\hat{c}(\hat{m})$. This implies, among other things, a modification of Eq. (7), that now becomes

$$p_{\hat{m}}(\hat{m}) = \hat{c}(\hat{m}) \int_{-\infty}^{\infty} \hat{p}_\epsilon(\hat{m}|m) p_m(m) dm . \quad (20)$$

This equation is valid for detected stars, i.e. if $\hat{m} \neq \text{null}$. Since $\hat{c}(\hat{m})$ represents the probability that a star with *measured* magnitude \hat{m} be detected, we evaluate the probability that the star is *not* detected as

$$p_{\hat{m}}(\text{null}) = \int_{-\infty}^{\infty} d\hat{m} [1 - \hat{c}(\hat{m})] \int_{-\infty}^{\infty} dm \hat{p}_\epsilon(\hat{m}|m) p_m(m) . \quad (21)$$

An important point to observe here is the fact that we have modified Eq. (7) into Eqs. (20) and (21) by defining *two* functions, the new error distribution $\hat{p}_\epsilon(\hat{m}|m)$ and the new completeness function $\hat{c}(\hat{m})$, which replace, respectively, $p_\epsilon(\hat{m}|m)$ and $c(m)$. Indeed, both these couples of functions are intimately related, and this was implicitly taken into account above by using a single distribution $p_\epsilon(\hat{m}|m)$ to describe both the photometric errors and the completeness of the observations. Interestingly, it is possible to find a relationship between the couple $\hat{p}_\epsilon(\hat{m}|m)$, $\hat{c}(\hat{m})$ and the couple $p_\epsilon(\hat{m}|m)$, $c(m)$ by requiring that, for any

reddened magnitude probability distribution $p_m(m)$, the observed magnitude probability distribution $p_{\hat{m}}(\hat{m})$ evaluated using Eq. (7), or using Eqs. (20) and (21), agrees. We find then

$$c(m)p_\epsilon(\hat{m}|m) = \hat{c}(\hat{m})\hat{p}_\epsilon(\hat{m}|m) , \quad (22)$$

$$1 - c(m) = \int_{-\infty}^{\infty} \hat{p}_\epsilon(\hat{m}|m) [1 - \hat{c}(\hat{m})] d\hat{m} . \quad (23)$$

Since $p_\epsilon(\hat{m}|m)$ is normalized to unity, if we integrate Eq. (22) over \hat{m} and substitute the result into Eq. (23), we find

$$\int_{-\infty}^{\infty} \hat{p}_\epsilon(\hat{m}|m) d\hat{m} = 1 , \quad (24)$$

i.e. \hat{p}_ϵ is also normalized to unity. In summary we have:

$$c(m) = \int_{-\infty}^{\infty} \hat{c}(\hat{m})\hat{p}_\epsilon(\hat{m}|m) d\hat{m} , \quad (25)$$

$$p_\epsilon(\hat{m}|m) = \frac{\hat{c}(\hat{m})\hat{p}_\epsilon(\hat{m}|m)}{\int_{-\infty}^{\infty} \hat{c}(\hat{m})\hat{p}_\epsilon(\hat{m}|m) d\hat{m}} . \quad (26)$$

Since Eq. (25) involves a convolution, it is in general not possible to invert it and express $\hat{c}(\hat{m})$ in terms of $c(m)$. Note that Eq. (26) is essentially Bayes’ theorem applied to $p_\epsilon(\hat{m}|m)$.

Although Eqs. (25) and (26) show that the description of the completeness $c(m)$ in terms of the true reddened magnitude m is more general than the description $\hat{c}(\hat{m})$ in terms of the observed magnitude \hat{m} , we argue that the latter is more practical to use:

- Operationally, the completeness function is usually evaluated from the *observed magnitudes* by comparing the expected number of stars in a given magnitude bin with the number of stars detected in the same bin.
- Often, it can be sensible to use some a posteriori cuts in the star catalog. For example, if we know that observations of faint stars are particularly unreliable, we can just discard all objects with observed magnitude \hat{m} larger than a given threshold m^{lim} . This cut could be easily described in terms of $\hat{c}(\hat{m})$:

$$\hat{c}(\hat{m}) = H(m^{\text{lim}} - \hat{m}) = \begin{cases} 1 & \text{if } \hat{m} \leq m^{\text{lim}} , \\ 0 & \text{otherwise} . \end{cases} \quad (27)$$

- The use of $\hat{c}(\hat{m})$ allows us to evaluate analytically $p_{\hat{m}}(\hat{m})$ for some special simple probability distributions. This point is particularly valuable for the practical applications that we will describe below in Sect. 4.

2.4. Multi-band probability distributions

So far we have focused our analysis on single-band measurements. In order to extend the discussion to observations carried out in different bands, we need to generalize the relevant probability distributions.

We denote $\mathbf{M} = \{M_1, M_2, \dots, M_\Lambda\}$ the magnitudes of a star in Λ different bands; in general, we use bold symbols such as \mathbf{k} to indicate quantities that have different values in the various bands. We will treat these as vector quantities; in this way, e.g., the generalization of Eq. (1) can be written as

$$\hat{\mathbf{m}}_n = \mathbf{M}_n + 5 \log_{10} D_n - 5 + \mathbf{k}A_V + \boldsymbol{\epsilon}_n = \mathbf{m}_n + \boldsymbol{\epsilon}_n . \quad (28)$$

Since we are now working with observations in different bands, we need also to generalize $\rho(M, D)$. We define thus $\rho(\mathbf{M}, D)$, the probability to have a star with absolute magnitudes \mathbf{M} at distance D from us. Using this distribution we can rewrite Eq. (6) as

$$p_m(\mathbf{m}) = \int_0^\infty dD D^2 \int_0^\infty dA_V p_{A_V}(A_V|D) \times \rho(\mathbf{m} - 5 \log_{10} D + 5 - kA_V, D). \quad (29)$$

The generalization of Eq. (7) is also simple. Since photometric measurements in different bands are independent, we have to perform Λ different convolutions with the measurement error probability distributions $\{p_{\epsilon_\lambda}(\hat{m}_\lambda|m_\lambda)\}$ corresponding to the various bands:

$$p_{\hat{\mathbf{m}}}(\hat{\mathbf{m}}) = \int d^\Lambda m p_m(\mathbf{m}) \prod_{\lambda=1}^\Lambda p_{\epsilon_\lambda}(\hat{m}_\lambda|m_\lambda). \quad (30)$$

Note that we consider a star detected if it is detected in *at least* one band.

The integral of $p_{\hat{\mathbf{m}}}(\hat{\mathbf{m}})$ over all admissible values of $\hat{\mathbf{m}}$ gives the expected local density of stars σ :

$$\sigma = \int_{(\mathbb{R} \cup \{\text{null}\})^\Lambda \setminus \{\text{null}\}} p_{\hat{\mathbf{m}}}(\hat{\mathbf{m}}) d^\Lambda \hat{\mathbf{m}}. \quad (31)$$

Since σ gives the normalization of $p_{\hat{\mathbf{m}}}(\hat{\mathbf{m}})$, the conditional probability that a star with magnitudes $\hat{\mathbf{m}}$ be observed given the fact that the star is detected is $p_{\hat{\mathbf{m}}}(\hat{\mathbf{m}})/\sigma$.

When using the alternate completeness functions $\{\hat{c}_\lambda(\hat{m}_\lambda)\}$ expressed in terms of the observed magnitudes $\{\hat{m}_\lambda\}$, Eq. (30) can be rewritten as

$$p_{\hat{\mathbf{m}}}(\hat{\mathbf{m}}) = \int d^\Lambda \hat{m}' \prod_{\lambda=1}^\Lambda \left[\delta(\hat{m}_\lambda - \hat{m}'_\lambda) \hat{c}_\lambda(\hat{m}'_\lambda) + \delta(\hat{m}_\lambda - \text{null}) (1 - \hat{c}_\lambda(\hat{m}'_\lambda)) \right] \times \int d^\Lambda m p_m(\mathbf{m}) \prod_{\lambda=1}^\Lambda p_{\epsilon_\lambda}(\hat{m}'_\lambda|m_\lambda). \quad (32)$$

The combination of delta distributions inside the brackets in this equation ensures that $\hat{m}'_\lambda = \hat{m}_\lambda$ if $\hat{m}_\lambda \neq \text{null}$, and that we integrate over $(1 - \hat{c}_\lambda(\hat{m}'_\lambda))$ (the probability of *not* detecting the star) if $\hat{m}_\lambda = \text{null}$; the last integration is the usual convolution with the measurement errors. Similarly, Eq. (31) can also be written as

$$\sigma = \int d^\Lambda \hat{m} \left[1 - \prod_{\lambda=1}^\Lambda (1 - \hat{c}_\lambda(\hat{m}_\lambda)) \right] \times \int d^\Lambda m p_m(\mathbf{m}) \prod_{\lambda=1}^\Lambda p_{\epsilon_\lambda}(\hat{m}_\lambda|m_\lambda). \quad (33)$$

2.5. Further simplifications

As for the single-band case, we consider a number of realistic and useful simplifications that will allow us to take advantage of the formalism introduced above in practical cases.

First, we still take $\rho(\mathbf{M}, D)$ to be separable into $\rho(\mathbf{M}, D) = p_M(\mathbf{M})\rho(D)$. Furthermore, we adopt for $p_M(\mathbf{M})$ a simple functional form, which is sufficiently accurate for our purposes. Since we know that $p_m(m) \propto 10^{\alpha m}$ with α approximately independent of the band, and since stars appear to have a limited scatter in their NIR colors, we write

$$p_M(\mathbf{M}) = \nu 10^{\alpha M_1} \times \exp \left[-\frac{(M_a - M_1 - \chi_a)C_{ab}^{-1}(M_b - M_1 - \chi_b)}{2} \right], \quad (34)$$

where we have used Einstein's convection on repeated indexes. In other words, we suppose that M_1 is exponentially distributed, and that star colors are distributed as Gaussian random variables with averages $\langle M_a - M_1 \rangle = \chi_a$; C represents the covariance of these colors. This distribution can be rewritten in a more convenient way as (see also below Sect. 4.2)

$$p_M(\mathbf{M}) = \exp \left[-\frac{\mathbf{M}^T \mathbf{P} \mathbf{M} + 2\mathbf{Q}^T \mathbf{M} + R}{2} \right]. \quad (35)$$

The form (34) for $p_M(\mathbf{M})$ is particularly convenient for several reasons. First, we can greatly simplify Eq. (29) and obtain a result similar to Eq. (10):

$$p_m(\mathbf{m}) = a \int_0^\infty dA_V p_{A_V}(A_V) p_M(\mathbf{m} - kA_V), \quad (36)$$

where $p_{A_V}(A_V)$ and a are still given by Eqs. (11) and (12). Again, we observe that the fact that $p_m(\mathbf{m})$ depends only on the distance-weighted $p_{A_V}(A_V)$, implies that $p_{A_V}(A_V|D)$ is not an observable. Moreover, the forms (34) and (35) are unmodified by a reddening $\mathbf{M} \mapsto \mathbf{m} = \mathbf{M} + 5 \log_{10} D - 5 + kA_V$ and by a convolution with Gaussian photometric errors (see below Sect. 4.2).

3. Extinction measurements

We will describe in this section the star counts and color excess methods in order to better describe the advantages and limitations of them.

3.1. Star counts

In the 18th century the English astronomer William Herschel noted that some regions presented few stars and, following Newton's idea of a perfectly transparent space, interpreted these "holes in the sky" as a real lack of stars. This misconception survived the discovery of individual dark clouds (Barnard 1919) and had serious consequences on Shapley's calibration of the distance scale for Cepheids. The "discovery" of dust in our Galaxy took place only in 1930 when Trumpler showed its importance in dimming the light coming from distant open clusters. Finally, in recent decades dust has no longer been seen only as annoying "fog" obscuring the light of background sources, and has been shown to have a tremendous impact on the evolution of galaxies and on the formation of stars and stellar systems (see Li & Greenberg 2003 for a detailed historical review).

It has long been recognized (Wolf 1923) that measurements of the local density of stars in different regions of the sky can be used to map the extinction. The original technique consisted in comparing the number of stars in magnitude bins in regions subject to extinction with the number of stars in regions where the extinction is (supposedly) negligible. This technique was then improved by Bok (1956) which suggested to use counts up to a limiting magnitude to reduce the error. In the past, the star count technique was mainly applied to optical data (typically visually inspected Smith plates; e.g. Dickman 1978; Mattila 1986; Andreazza & Vilas-Boas 1996). In the last decade, however, near-infrared digital data have been available, and the star count technique has been finally applied to NIR observations (e.g. Cambresy et al. 1997). In this respect, a key role has been played by large NIR surveys such as the Two Micron All Sky Survey (2MASS; Kleinmann et al. 1994) and the Deep NIR southern Sky Survey (DENIS; Epchtein et al. 1997).

The star count method is easily described using our notation. We first note that, as for multi-band probability distributions, the integral of $p_{\hat{m}}(\hat{m})$ over \hat{m} gives the local density of stars σ (cf. Eq. (31)):

$$\begin{aligned} \sigma \equiv \langle \sigma \rangle &= \int p_{\hat{m}}(\hat{m}) d\hat{m} \\ &= \int d\hat{m} \hat{c}(\hat{m}) \int dm p_m(m) p_\epsilon(\hat{m}|m). \end{aligned} \quad (37)$$

Inserting here Eq. (10), we immediately obtain

$$\sigma = \sigma^{(0)} \int dA_V p_{A_V}(A_V) 10^{-\alpha k_\lambda A_V}, \quad (38)$$

where $\sigma^{(0)}$ is the average density expected in absence of extinction:

$$\sigma^{(0)} = \alpha \nu_\lambda \int d\hat{m} \hat{c}(\hat{m}) \int dm 10^{\alpha m} p_\epsilon(\hat{m}|m). \quad (39)$$

Equation (38) shows that the ratio $\sigma/\sigma^{(0)}$ between the average densities expected in presence and in absence of extinction is simply related to $p_{A_V}(A_V)$.

Clearly, a single measurement of $\sigma/\sigma^{(0)}$ cannot be used to derive the distribution $p_{A_V}(A_V)$. However, in the simplest case where all stars are background to a thin cloud, so that $p_{A_V}(A'_V) = \delta(A_V - A'_V)$, we find the classical relation

$$\frac{\sigma}{\sigma^{(0)}} = 10^{-\alpha k_\lambda A_V}, \quad (40)$$

which can be immediately inverted to obtain A_V (Bok 1956). Operationally, both densities σ and $\sigma^{(0)}$ are measured by dividing the number of stars observed against the cloud and in a control field by the angular size of the regions considered. Hence, the measurement errors on these quantities are due to the randomness on the local number of stars. If we ignore the correlation on the star positions, and thus assume that these are a homogeneous Poisson process (see, e.g., Cressie 1993), then the local number of stars follows a simple Poisson distribution.

Although the estimator (40) can be applied to thin clouds only, in principle more general situations can also be handled by taking advantage of the particular form of Eq. (38). We first observe that the integral appearing in the r.h.s. of this equation

Table 1. The extinction law in different infrared bands (taken from Rieke & Lebofsky 1985 and van de Hulst 1946).

λ	k_λ (Rieke & Lebofsky 1985)	k_λ (van de Hulst 1946)
<i>J</i>	0.282	0.246
<i>H</i>	0.175	0.155
<i>K</i>	0.112	0.089
<i>L</i>	0.058	0.045
<i>M</i>	0.023	0.033
<i>N</i>	0.052	0.013

is reminiscent of a Laplace's transform. Given a real function $f(x)$ defined for non-negative values x , its Laplace's transform is defined as

$$\mathcal{L}[f](k) = \int_0^\infty f(x) \exp(-kx) dx. \quad (41)$$

An important property of Laplace's transform is that it can be inverted: in other words, it is possible to obtain $f(x)$ provided one knows $\mathcal{L}[f](k)$ for any non-negative k . Using the notation just introduced, we can rewrite Eq. (38) as

$$\sigma = \sigma^{(0)} \mathcal{L}[p_{A_V}](\alpha k_\lambda \ln 10). \quad (42)$$

Hence, if we were able to measure the ratio $\sigma/\sigma^{(0)}$ for all positive values of the product αk_λ , we could in principle derive p_{A_V} . As indicated above, the constant $\alpha \simeq 0.34$ is approximately independent of the band considered; however, k_λ strongly depends on the band used to carry out the observations (see Table 1). Hence, the product αk_λ that appears in Eq. (42) can be varied among a *limited* set of values. In conclusion, although complete knowledge of p_{A_V} is clearly impossible, we can still use multiband observations to investigate p_{A_V} in more complex situations than the one considered in Eq. (40).

Let us consider, for example, the case of the thin cloud approximation described in Eq. (17). Since, in general, the fraction f of foreground stars is not known, we want to estimate both A_V and f in a given patch of the sky. (Although f can be taken to be constant in many cases, changes on this quantity have to be expected for different regions of large cloud complexes because of the geometry of the cloud and of changes of star densities due to the Galactic structure.) For this, we insert Eq. (17) in Eq. (38), thus obtaining a simple generalization of Eq. (40):

$$\frac{\sigma}{\sigma^{(0)}} = f + (1 - f) 10^{-\alpha k_\lambda A_V}. \quad (43)$$

Hence, we can in principle use two different measurements of $\bar{\sigma}$ in two different bands to deduce both f and A_V on the region considered. In practice, it is preferable to follow the approach described below in Sect. 4.

It is interesting to investigate in more detail the estimator derived from Eq. (43), i.e.

$$\hat{A}_V = -\frac{1}{\alpha k_\lambda} \log_{10} \left[\frac{N - \mathcal{A} \sigma^{(0)} f}{\mathcal{A} \sigma^{(0)} (1 - f)} \right], \quad (44)$$

where N is the number of stars found in the region investigated, and \mathcal{A} is its area. Assuming that $\sigma^{(0)}$ is measured without significant errors (this is possible by using a large control field),

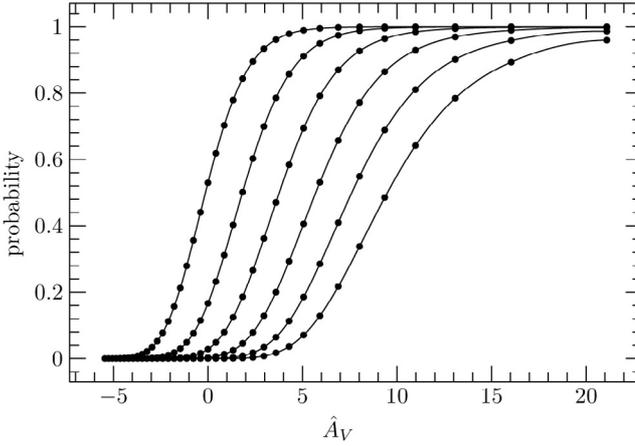


Fig. 1. The cumulative probability distribution for \hat{A}_V . The various curves are relative to a true extinction $A_V = [0, 2, 4, 6, 8, 10]$ (from the left to the right); the average values measured are instead $\langle \hat{A}_V \rangle = [0.24, 2.33, 4.47, 6.64, 8.79, 10.76]$ (the scatters around these values ranges from 1.9 to 4.3). In all cases we assumed $\alpha = 0.34$, $k_\lambda = 0.175$, $f = 0.1$, and $\mathcal{A}\sigma^{(0)} = 20$. Along the curves the dots mark the actual possible values of A_V that can be measured for different values of the observed number of stars N .

we can deduce the expected error on \hat{A}_V by noting that N follows a Poisson distribution with average (cf. Eq. (43))

$$\langle N \rangle = \mathcal{A}\sigma^{(0)} [f + (1-f)10^{-\alpha k_\lambda A_V}]. \quad (45)$$

Using a first order approximation (valid for small relative errors) we find

$$\text{Var}(\hat{A}_V) = \left(\frac{1}{\alpha k_\lambda \ln 10} \cdot \frac{1}{\langle N \rangle - \mathcal{A}f\sigma^{(0)}} \right)^2 \langle N \rangle. \quad (46)$$

In case of vanishing f (i.e., if all stars are background to the cloud), this expression takes a simpler form

$$\text{Err}(\hat{A}_V) = \sqrt{\text{Var}(\hat{A}_V)} = \frac{1}{\alpha k_\lambda \ln 10} \frac{1}{\sqrt{\langle N \rangle}} = \frac{11.4}{\sqrt{\langle N \rangle}}, \quad (47)$$

where the last expression is valid for the K band (cf. Table 1).

Further statistical properties of the estimator (44) can be better evaluated using numerical methods. Figure 1 shows, as an example, the expected cumulative probability distributions for \hat{A}_V for some typical cases. As expected, the cumulative distributions reach 0.5 around the true value for A_V , but the expected measurement errors can be very large, especially at high column densities. The numerical analysis of Eq. (44) also shows that the estimator is significantly biased. In general, in the limit $\sigma\mathcal{A} \gg 1$, the estimator is biased toward large values of \hat{A}_V , i.e. $\langle \hat{A}_V \rangle > A_V$; the opposite happens for $\sigma\mathcal{A} \sim 1$ (see Fig. 2). This change in behavior can be understood by observing that, if A_V is large, then $\mathcal{A}\sigma$ is small and thus we simply do not have enough background stars to probe high column densities.

Interestingly, Eq. (46) is a reasonably good approximation of the true variance of \hat{A}_V . Typically, this approximation slightly underestimates the true variance of \hat{A}_V , but again the opposite happens for low values of $\sigma\mathcal{A}$.

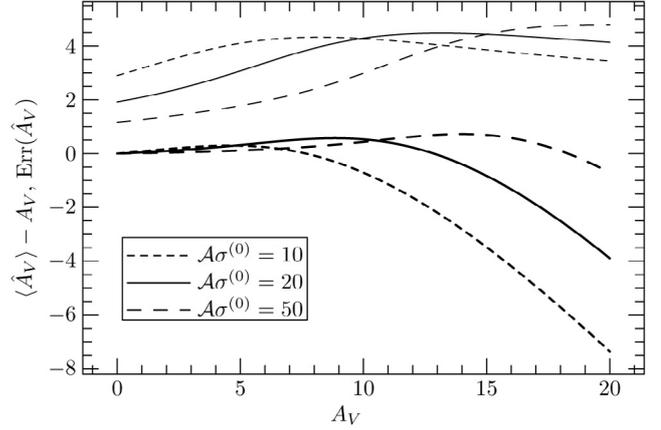


Fig. 2. The bias (thick lines) and error (thin lines) on \hat{A}_V evaluated using Eq. (44) for different values of $\mathcal{A}\sigma^{(0)}$. For this figure we used the same parameters as in Fig. 1, i.e. $\alpha = 0.34$, $k_\lambda = 0.175$, and $f = 0.1$.

3.2. Near-infrared color excess (NICE and NICER)

The dust present in molecular clouds produces different amounts of obscuration in different color bands (see Table 1), so that background stars appear reddened. Hence, the color excess in the red part of the spectrum of background stars can be used to measure the extinction along the line of sight.

Historically, the color excess technique has been impeded by the limited instrumental sensitivity to small regions (Jones et al. 1980; Frerking et al. 1982; Jones et al. 1984). However, in the early 1990s, with the advent of near-infrared arrays, it became possible to accurately measure the NIR magnitudes of many stars from single pointing observations. This new technology was first exploited by Lada et al. (1994), and since then has been successfully applied to many molecular clouds (e.g. Horner et al. 1997; Lada et al. 1999; Alves et al. 2001).

Lada’s technique (called “near-infrared color excess” or NICE) is based on measurements of a NIR color (e.g., $H - K$) of many stars. Since stars have relatively well defined colors in the infrared, a significant intervening extinction can be detected as a reddening. Note that a key point of the NICE method (and of similar methods based on the reddening of stars; see, e.g., Schultheis et al. 1999) is the assumption that all stars belong to a homogeneous population.

The NICE method can be quantitatively described using the notation introduced so far. In particular, from Eq. (29) we have

$$p_m(\mathbf{m}) = p_m^{(0)}(\mathbf{m} - \mathbf{k}A_V), \quad (48)$$

where we have denoted $p_m^{(0)}(\mathbf{m})$ the l.h.s. of Eq. (29) for $A_V = 0$. Naively, Eq. (48) can be used to obtain a simple estimate of the intervening infrared extinction A_V . For example, we find

$$\langle \mathbf{m} \rangle = \langle \mathbf{m} \rangle^{(0)} + \mathbf{k}A_V, \quad (49)$$

where again we used the superscript (0) to denote quantities measured in a control region where $A_V = 0$. Equation (49) is better rewritten in terms of star colors. Calling $\chi_\lambda = m_\lambda - m_1$ and $\kappa_\lambda = k_\lambda - k_1$, we have

$$\langle \chi \rangle = \langle \chi \rangle^{(0)} + \kappa A_V. \quad (50)$$

This equation, applied to a single color, is essentially the NICE technique. More precisely, this technique uses the simple average of a set of N angularly close stars to evaluate the column density:

$$\hat{A}_V = \frac{1}{N} \sum_{n=1}^N \frac{\hat{\chi}_n - \bar{\chi}^{(0)}}{\kappa}, \quad (51)$$

where $\bar{\chi}^{(0)}$ is an estimate of $\langle \chi \rangle^{(0)}$ (obtained, e.g., by measuring the star colors on a control field where presumably $A_V \simeq 0$). As an example, consider the NICE method applied to the $\chi = H - K$ color. Since stars have a typical scatter of 0.09 mag in this color, we expect an error on \hat{A}_V of $\text{Err}(\hat{A}_V) \simeq 1.4/\sqrt{N}$. Hence, even in the presence of significant photometric errors, the NICE method gives significantly more accurate results than the star count method (cf. Eq. (47)).

As shown by Lombardi & Alves (2001), one can indeed take full advantage of observations carried out in different bands to obtain more accurate column density measurements. The improved technique, called NICER (NICE Revised) optimally balances the information from different bands and different stars. As a by-product of the analysis, NICER also allows us to evaluate the expected error on the column density map, which is useful to estimate the significance on the detection of substructures and cores. The NICER technique can be described using the following simple argument. Equation (50) written above can be taken to be a system of $(\Lambda - 1)$ equations to be approximately solved for A_V , the approximation being made necessary because we can only measure $\langle \chi \rangle$ and $\langle \chi \rangle^{(0)}$ with limited accuracy. The “best” solution for A_V can be obtained by minimizing the chi-square quantity

$$\chi^2 = \sum_{n=1}^N [\hat{\chi}_n - \bar{\chi}^{(0)} - \kappa A_V]^T (C + E)^{-1} [\hat{\chi}_n - \bar{\chi}^{(0)} - \kappa A_V], \quad (52)$$

Consistently with the notation used above in Eq. (34), we called C the covariance matrix of the star colors; moreover, the symbol E was used to denote the covariance matrix of measurement errors (the two covariance matrices have to be added up in Eq. (52) in order to properly estimate the expected scatter on star colors). The best estimate of A_V , obtained by minimizing the χ^2 of Eq. (52), is precisely the NICER estimator.

Both the NICE and NICER estimators *appear* to be unbiased provided that: (i) there are no foreground stars and (ii) the measured colors $\hat{\chi}_n$ are unbiased estimates of the true colors χ_n .

In reality, even if the two conditions considered above are satisfied, both color excess methods can still be biased because of selection effects introduced by the completeness function. To understand this point let us make a simple example. Suppose that we carry out our observations in two bands, λ_1 and λ_2 , and that both completeness functions $c_\lambda(m_\lambda)$ are not vanishing only on a narrow magnitude range. In this case we would always have $\hat{\chi} \simeq \hat{\chi}^{(0)}$ (because a star is observed in both bands only if m_1 and m_2 are inside the narrow detection window), and thus we would always measure $\hat{A}_V \simeq 0$, independently of the real column density. Although unrealistic, this example shows that we must expect a bias for the NICE method; the argument for the NICER method is similar and leads to the same conclusion.

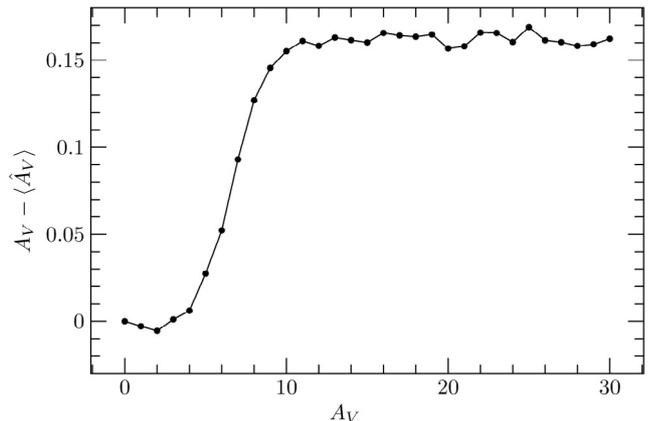


Fig. 3. The bias for the NICE method due to the different completeness at low and high column densities. This plot is evaluated by generating 100 000 stars for various values of A_V (marked as dots). Note that, as described in the text, for $A_V \simeq 6.7$ we observe a rapid increase in the bias due to the change between a selection in the K band and a selection in the H band (see Fig. 4). The small scale oscillations on the plot are due to numeric effects.

The bias of these two methods depends on the details of the probability distribution $p_M(\mathbf{M})$ and of the completeness functions $c_\lambda(m_\lambda)$. However, as an example, we evaluated the expected bias for various values of the column density and for the typical probability distributions that we expect for the 2MASS catalog. As shown in Fig. 3, the bias in the case considered appears to be limited, below 0.2 in A_V , and has a characteristic shape: it vanishes for $A_V = 0$, increases quickly for $A_V \sim 7$, and finally saturates for $A_V > 11$. This behavior has a simple explanation:

- Since the colors of unreddened stars are evaluated using a control field (where supposedly $A_V = 0$), the bias has to vanish for $A_V = 0$.
- The general trend of bias on \hat{A}_V can be understood with the help of Fig. 4. At low A_V , stars that are observed in the K band are almost certainly also observed on the H band, because of the values of the limiting magnitudes (approximately 14.9 in H and 14.3 in K) and of the average star colors ($\langle H - K \rangle = 0.18$). When, instead, the reddening is large, we have the opposite situation (because K is less affected by reddening than H). This different selection at different column densities is the source of the observed bias.
- Finally, at large column densities an asymptotic value is reached because now only the H band is used to select stars.

A more serious problem is related to the contamination by foreground stars, which can strongly bias our results in the direction of lower column densities. For regions with low extinction, where the expected number of foreground stars is small compared to the expected number of background stars, foreground star contamination is usually reduced by using a *median* estimator for \hat{A}_V , i.e. by averaging the individual column densities measured in the direction of each star with the median instead of the simple mean (e.g. Cambr esy et al. 2002; Lombardi & Alves 2001). As shown in Fig. 5, this simple technique is very effective and leads to *almost* unbiased results at

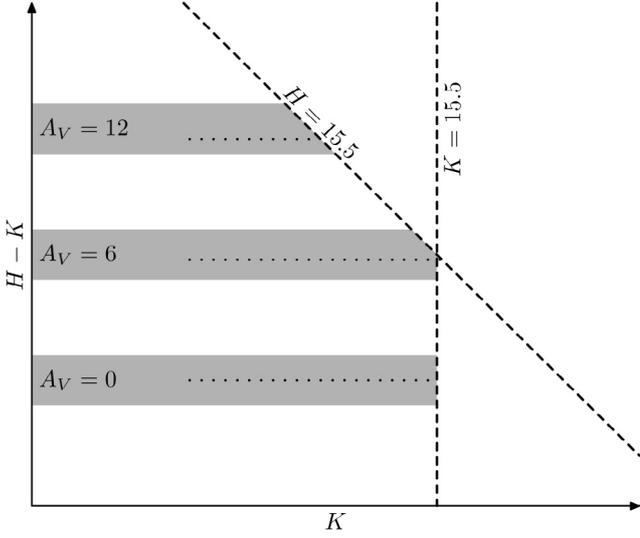


Fig. 4. A graphic explanation of the bias plotted in Fig. 3. At low column densities, we select stars mainly on their K magnitude, while at high column densities we mainly select using their H magnitude. As a result, the average intrinsic color (dotted line) of the observed stars (gray bands) changes toward lower $H - K$ values.

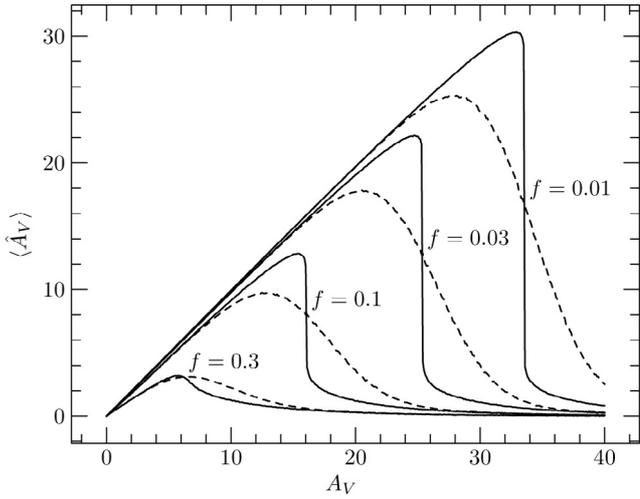


Fig. 5. The bias for the NICE method due to the contamination by foreground stars. Solid lines represent, for different fractions of foreground stars, the median of the distribution of the measured column densities. Dashed lines, instead show the expectation value for the median of \hat{A}_V obtained from the measurements of 15 reddening.

low column densities (see Appendix A for a statistical discussion on the median estimator). However, for a given fraction of foreground stars f (which, as described above, is evaluated in regions with negligible extinction) there is a threshold for A_V after which the cloud extinction makes the expected number of background stars smaller than the expected number of foreground stars. This threshold can be evaluated using Eq. (43) and is

$$A_V = \frac{1}{\alpha k_\lambda} \log_{10} \frac{1-f}{f}. \quad (53)$$

Because of the way the median estimator is defined, we observe in Fig. 5 an abrupt change on the measured column

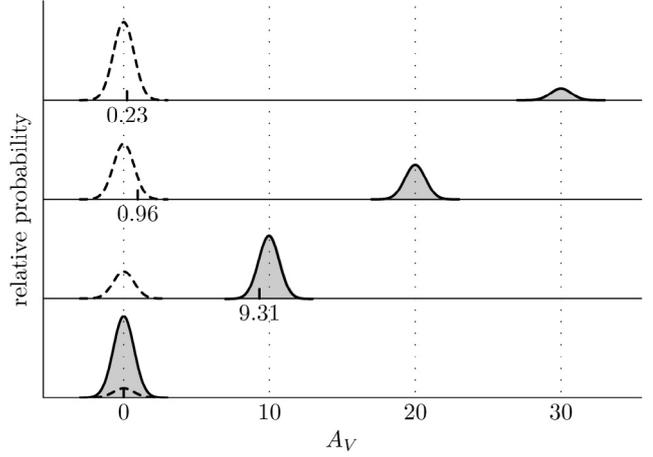


Fig. 6. The effect of a foreground star contamination $f = 0.1$ on the distribution of column densities for different values of A_V . From the bottom to the top we show, for $A_V \in \{0, 10, 20, 30\}$, both the distribution of background stars (solid plot, gray filled), and foreground stars (dashed plot). Both distributions are taken to be Gaussian, the first centered on A_V , and the second on 0. As the column density increases, the relative contribution of foreground stars also increases because of selection effects. At $A_V \approx 16$, they become predominant and the median of the whole distribution quickly moves toward the foreground distribution.

density \hat{A}_V close to this value (see also Fig. 6). Note that the relatively smooth transition observed in Fig. 5 (solid lines) is due to the intrinsic scatter in the star colors; indeed, in absence of any scatter, we would observe an “instantaneous” change from $\langle \hat{A}_V \rangle = A_V$ at low column densities to $\langle \hat{A}_V \rangle = 0$ at high column densities. In Fig. 5 we also plot the expectation value of a more interesting quantity, the median over $N = 15$ measured column densities. This quantity differs from the median over the whole distribution because of the statistical variations on the local number of foreground stars. In other words, since $N = 15$ is a relatively small number, in different realizations of our simulations we can have a significantly different number of foreground stars. For example, even at low column density, we can have a large fraction of foreground stars; similarly, even at very large column density, there is a finite probability to have the majority of stars in the background. As a result, the dashed line in Fig. 5 has a much smoother transition around the value given by Eq. (53). This is at the same time a good and a bad news. From one side, this means that we can be significantly biased for values of A_V smaller than the threshold value of Eq. (53); from the other side, this also means that we can still partially make use of the median estimator at relatively large column densities.

For regions with very high column density, it is normally quite easy to identify and remove foreground stars because of their anomalous colors with respect to the reddened background stars (see, e.g., Alves et al. 2001). Some authors (e.g. Cambr sy et al. 2002) make use of this information to also remove stars in less dense regions using the following strategy. The angular density σ_f of foreground stars is determined using dense regions (where the foreground/background identification is easy); then, for any region of the cloud, the k bluer stars

are thrown away, where $N_f = \sigma_f \mathcal{A}$ is the expected number of foreground stars in the analyzed region (deduced from the foreground density measured in the dense regions). This technique is quite simple and reasonably effective, but unfortunately introduces a bias at low extinctions. Consider, indeed, the limiting case where we have a negligible extinction $A_V \simeq 0$. The distinction between foreground and background stars is in this situation ambiguous, and thus the N_f bluer objects will likely include also some background stars. Hence, the results will be biased toward higher column density. The exact evaluation of the bias is non trivial, but can be carried out using the techniques described in Appendix A. Here we report only an approximated result valid for $A_V \simeq 0$ (see Eq. (A.15)),

$$\langle \hat{A}_V \rangle \simeq \sqrt{2\pi} \text{Err}(\hat{A}_V) \frac{N_f}{2N}, \quad (54)$$

where $\text{Err}(\hat{A}_V)$ is the average error of the measured extinction from a single star. Note that the result given in Eq. (54) can be taken to be an upper limit for the bias, since we expect this to decrease at high column densities, where the identification of foreground stars is secure.

4. A maximum-likelihood approach

From the discussion above, it is clear that both the star count and the NICE(R) methods can produce unsatisfactory results. On one hand, the star count technique has a low signal-to-noise ratio and produces significantly biased results at high column densities (see Fig. 2). On the other hand, the color excess technique is more sensitive and has a smaller bias, but can be severely affected by the contamination of foreground stars, especially for large values of A_V .

As pointed out by Cambr esy et al. (2002), we can take advantage of the different behavior of the star count and the color excess techniques in the various column density regimes to partially solve the problem of the contamination by foreground stars. In particular, Cambr esy et al. note that it is better to use the color excess method at low column densities because of its higher sensitivity, and to switch to the star count method at very large column densities where the NICE method is completely unreliable. The optimal ‘‘turning point’’ where we need to change the technique can be evaluated empirically by comparing the individual results of the two methods at different column densities.

The solution proposed by Cambr esy et al. has the significant advantage of being relatively simple to implement and reasonably effective, but clearly is suboptimal in many respects:

- The choice of the ‘‘turning point’’ and of the matching strategy is to some extent arbitrary.
- The overall estimate of A_V still remains significantly biased at high column densities because at large A_V the star counts are used (and this method has a non-negligible bias, see Fig. 2).
- The density of foreground stars must still be determined separately and is taken to be constant on the whole field.

Using the theoretical framework developed so far, it is possible to design and implement an efficient maximum-likelihood approach to the problem.

4.1. Likelihood

Suppose again that in a region of the sky of area \mathcal{A} we observe in various bands N stars with magnitudes $\{\hat{m}_n\}$. Assuming that the area of the sky is small enough so that there are no significant changes on the relevant parameters of the problem (such as the unreddened density $\sigma^{(0)}$, local expected density σ , or the reddening probability distribution $p_{A_V}(A_V)$), then we can easily evaluate the joint probability distribution for such a star configuration. First, we note that the number of stars inside the region follows a Poisson distribution with average $\mathcal{A}\sigma$:

$$p_N(N) = e^{-\mathcal{A}\sigma} \frac{(\mathcal{A}\sigma)^N}{N!}. \quad (55)$$

The joint star probability distribution, i.e. the likelihood, is given by

$$\begin{aligned} L(\{\hat{m}_n\}) &= p_N(N) \prod_{n=1}^N \frac{p_{\hat{m}_n}(\hat{m}_n)}{\sigma} \\ &= \frac{e^{-\mathcal{A}\sigma} \mathcal{A}^N}{N!} \prod_{n=1}^N p_{\hat{m}_n}(\hat{m}_n). \end{aligned} \quad (56)$$

Note that L depends on unknown quantities, such as $p_{A_V}(A_V)$, through $p_N(N)$, $p_{\hat{m}_n}(\hat{m}_n)$, and σ : hence, assuming that there is no *prior* for these unknown quantities, we can obtain an estimate of them by maximizing L or, equivalently, $\ln L$. In the following subsections we will investigate in more detail this maximum-likelihood estimator.

4.2. Implementation

We implemented the maximum-likelihood estimator using the simplification described above. In particular, we used the forms (34) and (35) for the source magnitude probability distribution $p_M(\mathbf{M})$. A simple calculation gives the following relationships between the coefficients of Eqs. (34) and (35):

$$\begin{aligned} P_{ab} &= C_{ab}^{-1} - \delta_{1a} \sum_{a'} C_{a'b}^{-1} - \delta_{1b} \sum_{b'} C_{ab'}^{-1} \\ &\quad + \delta_{1a} \delta_{1b} \sum_{a',b'} C_{a'b'}^{-1}, \end{aligned} \quad (57)$$

$$Q_a = -\delta_{a1} \alpha \ln 10 - C_{ab}^{-1} \chi_b + \delta_{1a} \sum_{a'} C_{a'b}^{-1} \chi_b, \quad (58)$$

$$R = -2 \ln n + C_{ab}^{-1} \chi_a \chi_b. \quad (59)$$

One of the advantages of the quadratic expression (35) is that we can write the effects of reddening as a simple transformation of the three quantities P , Q , and R . In particular, the transformation $\mathbf{M} \mapsto \mathbf{M} + \mathbf{k}A_V$ can be rewritten as

$$\begin{aligned} P &\mapsto P, & Q &\mapsto Q - A_V P \mathbf{k}, \\ R &\mapsto R - 2A_V \mathbf{Q}^T \mathbf{k} + A_V^2 \mathbf{k}^T P \mathbf{k}. \end{aligned} \quad (60)$$

Hence, reddening does not change the functional form of the probability distribution (35) but only the three parameters involved.

For the following discussion, it is also convenient to introduce a new vector $\boldsymbol{\mu} \equiv (\mathbf{M}, 1) = (M_1, M_2, \dots, M_\Lambda, 1)$, and a $(\Lambda + 1) \times (\Lambda + 1)$ matrix S defined as

$$S^{-1} = \begin{pmatrix} P & \mathbf{Q} \\ \mathbf{Q}^T & R \end{pmatrix} = \begin{pmatrix} P_{11} & \dots & P_{1\Lambda} & Q_1 \\ \vdots & \ddots & \vdots & \vdots \\ P_{\Lambda 1} & \dots & P_{\Lambda\Lambda} & Q_\Lambda \\ Q_1 & \dots & Q_\Lambda & R \end{pmatrix}. \quad (61)$$

Then we can write the quadratic form appearing in Eq. (35) simply as

$$p_M(\mathbf{M}) = \exp\left[-\frac{\boldsymbol{\mu}^T S^{-1} \boldsymbol{\mu}}{2}\right]. \quad (62)$$

More importantly, the action of measurement errors can be described as simple transformations of S , provided that the measurement error on each band can be described as a simple Gaussian with vanishing average and variance \mathbf{v} . In other words, assuming that the errors can be represented as a convolution with the Gaussian kernel

$$p_\epsilon(\boldsymbol{\epsilon}) = \frac{1}{(2\pi)^{\Lambda/2} \prod_{\lambda=1}^{\Lambda} v_\lambda} \exp\left[-\sum_{\lambda=1}^{\Lambda} \frac{\epsilon_\lambda^2}{2v_\lambda^2}\right], \quad (63)$$

then the convolution $p_M * \Sigma$ can be written as

$$(p_M * \Sigma)(\boldsymbol{\mu}) = \sqrt{\frac{\det S}{\det T}} \exp\left[-\frac{\boldsymbol{\mu}^T T^{-1} \boldsymbol{\mu}}{2}\right], \quad (64)$$

where $T = S + \text{diag}(v_1^2, v_2^2, \dots, v_\Lambda^2, 0)$.

In conclusion, the simple implementation of the maximum-likelihood method considered here can be summarized in the following items:

1. The quantities P , \mathbf{Q} , and R appearing in Eq. (35) are determined in a control field free from any extinction. Similarly, for each band the photometric errors, i.e. the functions $p_{\epsilon_\lambda}(\hat{m}_\lambda | m_\lambda)$, are calculated.
2. A cloud model (for example, the thin cloud approximation) is chosen, and the relevant parameters (for example, A_V and f) are taken to be unknown and constant on the field.
3. For a given combination of the cloud parameters, the likelihood function (56) is evaluated using the various observed magnitudes. Note that the evaluation of the expected star density σ and of the star probability $p_{\hat{m}}(\hat{\mathbf{m}})$ when the star is undetected in one or more bands is non-trivial and requires integrations over the completeness functions (cf. Eq. (32)).
4. The last step is repeated for different values of the cloud parameters and the ones corresponding to the minimum of the likelihood function are taken as best estimates.
5. The local curvature of L (i.e., the matrix of its second derivatives) is used to estimate the errors on the cloud parameters.

Clearly, one key point here is the speed of the likelihood function, which needs to be evaluated several times in our maximum-likelihood approach. In our implementation, the likelihood function has been optimized by performing the relevant integrations (cf. point 3 above) using appropriate bounds.

Table 2. The common parameters used for all numerical simulations.

Par.	Value	Description	Ref. Eq.
α	0.34	Slope of number counts	(34)
χ_1	0.18	Average color $\langle H - K \rangle$	(34)
χ_2	0.82	Average color $\langle J - K \rangle$	(34)
C_{11}	0.0078	Variance $\langle (H - K - \chi_1)^2 \rangle$	(34)
C_{12}	0.0112	Covariance $\langle (J - K - \chi_1)(H - K - \chi_2) \rangle$	(34)
C_{22}	0.0375	Variance $\langle (J - K - \chi_2)^2 \rangle$	(34)
k_1	0.112	Reddening law in K band	(8)
k_2	0.175	Reddening law in H band	(8)
k_3	0.282	Reddening law in J band	(8)
m_1^{lim}	14.3	Limit magnitude in K band	(20) and (27)
m_2^{lim}	14.9	Limit magnitude in H band	(20) and (27)
m_3^{lim}	15.8	Limit magnitude in J band	(20) and (27)
v_1	0.05	Average photometric error $\langle (\hat{m}_1 - m_1)^2 \rangle^{1/2}$ on K band	(20)
v_2	0.05	Average photometric error $\langle (\hat{m}_2 - m_2)^2 \rangle^{1/2}$ on H band	(20)
v_3	0.05	Average photometric error $\langle (\hat{m}_3 - m_3)^2 \rangle^{1/2}$ on J band	(20)

In other words, when an integration of $p_{\hat{m}}(\hat{\mathbf{m}})$ was requested, we estimated the area in the magnitude space where this function was significantly different from zero, and performed the integral inside that area (as opposed to performing the integral over the whole parameter space). This optimization was found to have a significant impact on the overall speed of our implementation.

4.3. Simulations

The reliability and effectiveness of the maximum-likelihood approach were assessed through extensive numerical simulations. The simulations were designed to reproduce with reasonable accuracy the 2MASS near-infrared data. We simulated star observations in three bands, J , H , and K , and used the various parameters described in Table 2.

We initially considered an area of the sky \mathcal{A} and a thin cloud characterized by a fraction of foreground stars f (cf. Eq. (18)) and a reddening A_V . We randomly generated there stars inside this area using the following recipe:

1. We evaluated the expected local star density σ using Eqs. (31) and (32). We found that the needed integrations could be performed more efficiently using a Monte-Carlo technique.
2. We calculated the effective number of stars N by generating a random integer distributed according to a Poisson distribution with average $\mathcal{A}\sigma$.
3. For each of the N star we adopted the following procedure:
 - (a) We generated the unreddened magnitudes in the three bands according to Eq. (34).
 - (b) We then uniformly generated a random number in the range $[0, 1]$, and considered the star to be in the foreground if this number was smaller than f (defined according to Eq. (18)).

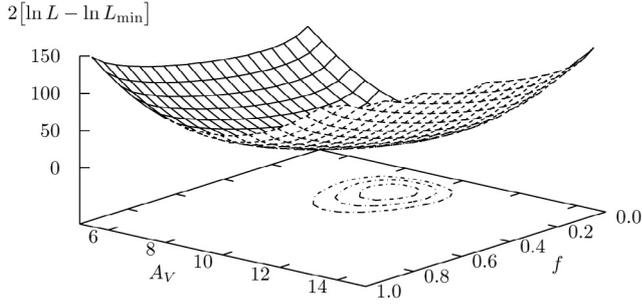


Fig. 7. Log-likelihood surface plot. The plot shows the logarithmic of the likelihood ratio as a function of the two unknown parameters A_V and f (the real values used for the simulation are 10 and 0.2 respectively). The simulation has been carried out using $\sigma_{\mathcal{A}} = 25$, but the actual number of stars generated were $N = 19$ (because of the Poisson noise on the number of stars). On the bottom we also plot contours corresponding to 68.2%, 95.5%, and 99.7% confidence level regions. Note that the likelihood is very smooth and has only a single well defined minimum.

- (c) Background stars were reddened by adding $k_{\lambda}A_V$ to each magnitude; the magnitudes of foreground stars were left unchanged.
- (d) We added photometric errors to all stars; these, for simplicity, were taken to be Gaussian distributed with standard deviation $v_{\lambda} = 0.05$ independent of the band and of the original magnitude.
- (e) For each band, we uniformly generated a random number in the range $[0, 1]$, and took the star to be detected in the band if this number was smaller than the completeness function $\hat{c}_{\lambda}(\hat{m}_{\lambda})$. In the simulations discussed here we used for simplicity Heaviside functions for the completeness functions.
- (f) We finally retained the star if it was detected in at least one band; otherwise, we repeated all steps above from point (a).

In summary, at the end of a single star generation we had for each star the three magnitudes in the bands J , H , and K (with possibly some magnitudes $\hat{m}_{\lambda} = \text{null}$) and the associated measurement errors v_{λ} .

We then used this dataset to test the reliability and efficiency of the maximum-likelihood estimator, and to compare it with the other column density estimators considered in Sect. 3. The maximum-likelihood method was implemented as described in Sect. 4.2, and was tested against the data generated as described in the items above.

Figure 7 shows the log-likelihood surface plot obtained in a typical simulation. The surface appears to be very smooth with a well defined minimum, an essential condition for the reliability of the maximum-likelihood approach. Moreover, the log-likelihood function is found to have approximately a parabolic shape, which further simplifies the interpretation of the results obtained. For example, this property allows us to use the likelihood ratio (see, e.g., Eadie et al. 1971) to draw confidence level regions on the parameter space (see the contours of Fig. 7).

Figure 8 represents the log-likelihood as a function of A_V for three different datasets. The figure was obtained by minimizing the log-likelihood function with respect to f for each

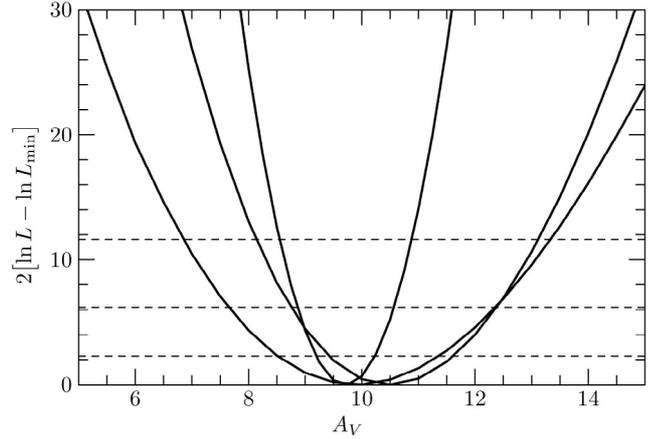


Fig. 8. The log-likelihood (ratio) as a function of A_V for different source densities, extremized for f . The largest curve is obtained using $\sigma_{\mathcal{A}} = 10$ (in this particular realization we had $N = 8$); the mid curve using $\sigma_{\mathcal{A}} = 20$ ($N = 12$), and the most peaked curve using $\sigma_{\mathcal{A}} = 40$ ($N = 35$). In all cases the true value of the column density was $A_V = 10$, and the fraction of foreground stars was $f = 0.2$. Note that all curves are very well approximated by parabolas. The intersections of the log-likelihood functions with the three dashed lines mark the 68.2%, 95.5%, and 99.7% confidence level regions.

value of A_V in the range $[5, 15]$, and by plotting the value of this function.

In order to assess more quantitatively the merits of the maximum-likelihood method, we compared the statistical properties of various column density estimators. In particular, we simulated a large number of “observations” using the technique described above, and we studied the distribution of the column densities estimated using the maximum-likelihood method, the NICE, and the NICER methods. Simulations were carried out using a thin cloud with true extinction in the range $A_V \in [0, 30]$ and with different values of the foreground fraction f . For the NICE and NICER estimators we used both the simple mean and the median of the individual extinction measured for each star; moreover, assuming that the density of foreground stars σ_f could be determined separately, we discarded in each simulation the $\sigma_f \mathcal{A}$ bluer stars, and used only the remaining (redder) stars in the analysis. Note that in some cases, for large values of A_V and relatively large values of f we had no usable star for the NICE and NICER analysis; in other words, all stars left after the foreground selection had only one band available (typically the K band). In this case we just obtained a lower limit on A_V by using the redder usable star (even if this star was originally considered to be in the foreground).

Figures 9 and 10 show, respectively, the bias and the total error obtained from the three methods considered here for $A_V \in [0, 30]$ and $f = 0$. From these plots it is evident that the maximum-likelihood estimator does not have any significant bias up to $A_V = 20$ and a very small one (of the order of 0.1) for larger column densities. Since for $f = 0$ we never have foreground stars, the bias of the NICE and NICER techniques does not change if we use a mean or a median estimator. Note also that the bias in Fig. 9 for these two methods is the one discussed in detail above (cf. Fig. 3). Regarding the total error,

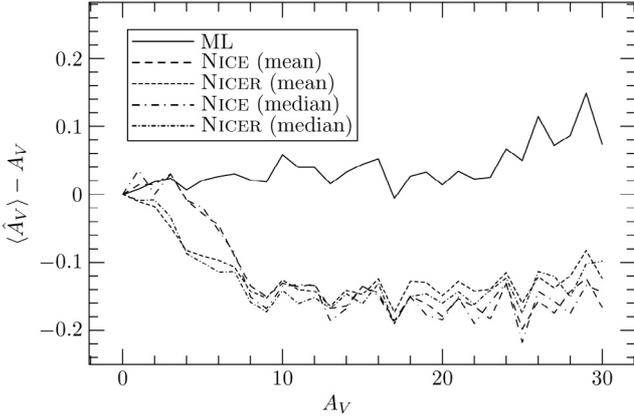


Fig. 9. The bias $\langle \hat{A}_V \rangle - A_V$ of various column density measurement methods. The bias is evaluated from the averages of 1 000 simulated fields with no foreground contribution ($f = 0$) and with average number of stars $\sigma\mathcal{A} = 25$. The maximum-likelihood method, marked in the legend as ML for brevity, has negligible bias, especially for $A_V < 20$.

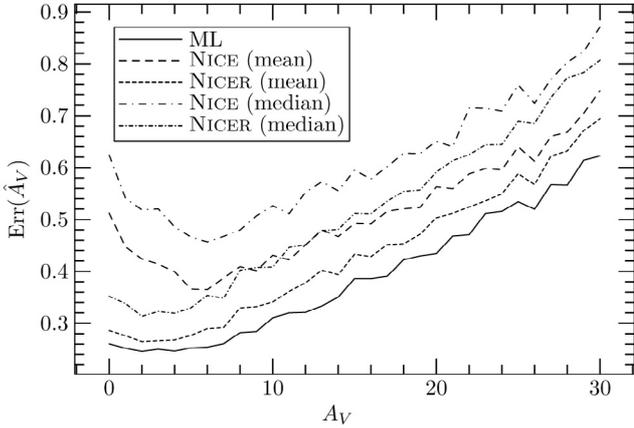


Fig. 10. The total error $\langle (\hat{A}_V - A_V)^2 \rangle^{1/2}$ of various column density measurement methods, evaluated as in Fig. 9.

we observe a steady increase of it for large values of A_V . This can be explained by noting that, although the average number of stars $\sigma\mathcal{A} = 25$ is kept constant for all column densities, when A_V is large most stars are only detected in the K band and thus do not provide reddening information. Figure 10 also shows that the maximum-likelihood method is clearly superior, although NICER (with the mean estimator) also performs well. As expected, both median estimators are more noisy than the simple mean (which, for $f = 0$, is optimal).

The situation changes quite dramatically when $f > 0$. Figures 11–16 show the bias and the total error of the various methods for increasing values of f . A careful study of these results can reveal several interesting characteristics of the NICE and NICER techniques.

Let us initially focus on the bias plots, shown in Figs. 11, 13, and 15. At low column densities, i.e. for $A_V < 10$, we find again the bias described in Sect. 3.2 and plotted in Fig. 3 (see also above Fig. 9 for the case $f = 0$). At median column densities, i.e. for $A_V \sim 20$ (or $A_V \sim 30$ for $f = 0.02$), the bias can be explained by the selection effect due to the correction of

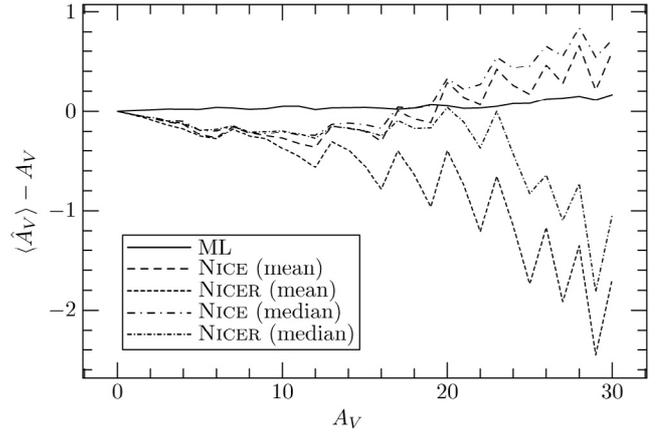


Fig. 11. As for Fig. 9, but with $f = 0.02$.

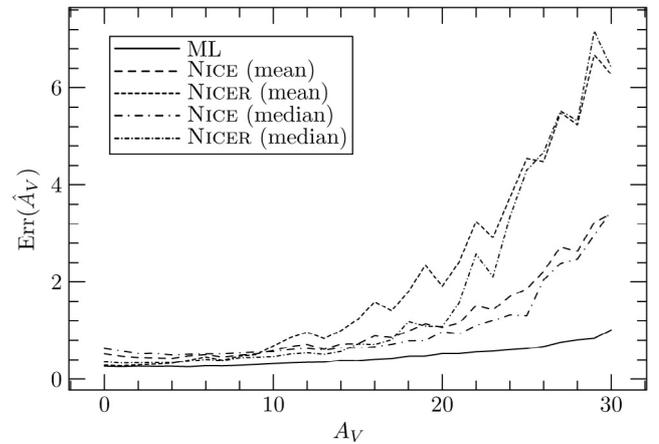


Fig. 12. As for Fig. 10, but with $f = 0.02$.

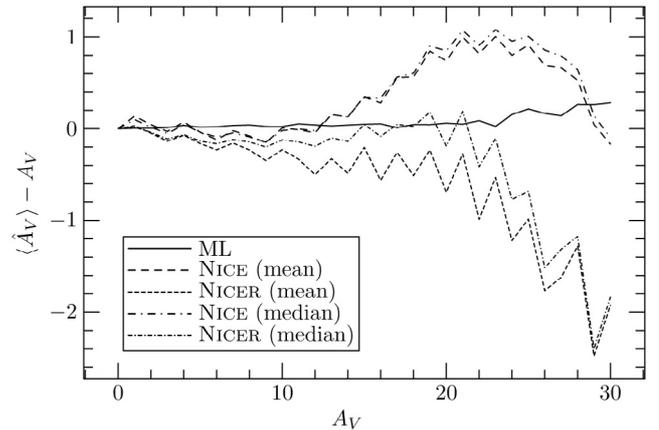


Fig. 13. As for Fig. 9, but with $f = 0.05$.

foreground stars. As explained above (see Eq. (54)), removing the $N_f = \sigma_f\mathcal{A}$ bluer stars introduces a systematic error on the extinction estimate. This bias is positive (i.e. toward larger extinctions) and can be as large as 1 mag in A_V . Finally, at high column densities ($A_V \sim 30$) both methods systematically underestimate the extinction. This is due to the heavy contamination by foreground stars present at these large values of reddening. To better understand this point, we note that the subtraction of the N_f bluer stars is a simplistic approximation because the

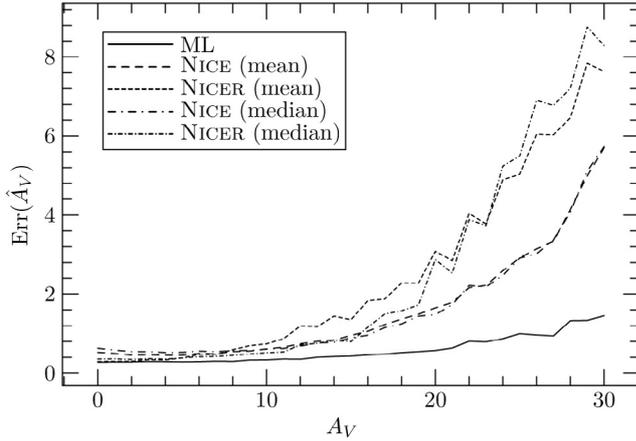


Fig. 14. As for Fig. 10, but with $f = 0.05$.

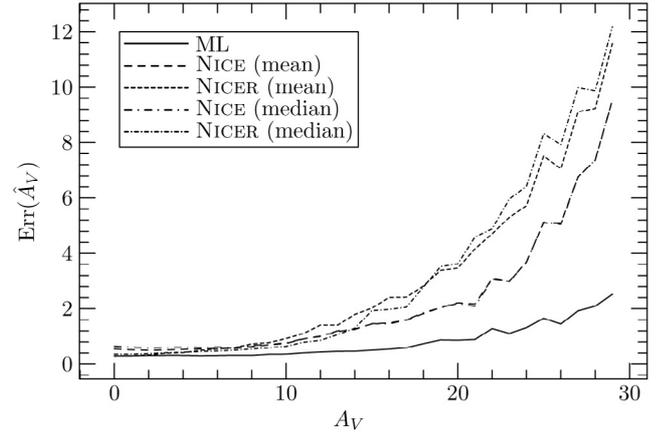


Fig. 16. As for Fig. 10, but with $f = 0.10$.

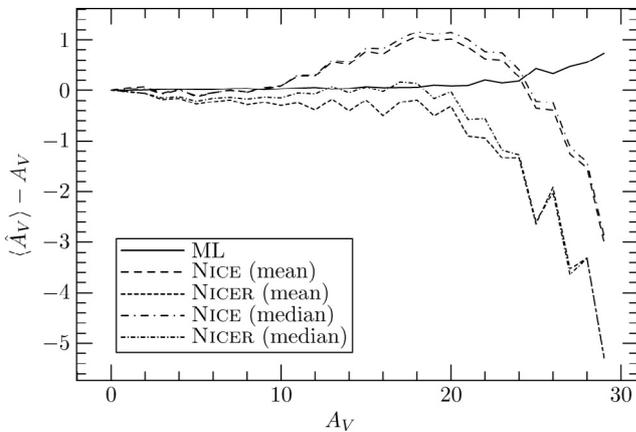


Fig. 15. As for Fig. 9, but with $f = 0.10$.

actual number of foreground stars is not fixed (it is a Poisson random variable with average N_f). Depending on the number of foreground stars with respect to N_f we can have three different situations:

- If the number of foreground stars is exactly N_f , then at high column densities no bias is introduced and the estimate of A_V is correctly performed;
- For a simulation with a number of foreground stars smaller than N_f , a small bias toward higher column densities is expected, because some of the bluer background stars are discarded;
- Finally, if the number of foreground stars is underestimated, a very large bias toward smaller column densities is introduced.

Clearly, for large values of A_V the third effect is expected to dominate. Indeed, Figs. 13 and 15 show that both the NICE and NICER methods significantly underestimate the reddening for large values of A_V . In theory, as mentioned above, when applying the NICE or NICER technique to high extinction regions, it should be relatively straightforward to identify foreground stars by their color, and thus the bias of these method could be smaller than suggested by the plot shown here. Note that apparently NICER presents a larger bias compared to NICE. This is due to the larger flexibility of the NICER method, which is

able to obtain a column density estimate when any two of the three bands are available. As a result, NICER is more affected by the contamination by foreground stars described in the items above. Indeed, our simulations show that if we force NICER to use only stars with observed H and K bands (i.e. essentially the same stars as the ones used by NICE), its bias and its noise are drastically reduced and become compatible with the ones of NICE.

Figures 12, 14, and 16 show that the total error of the NICE and NICER methods increases very rapidly at large column densities, where the contamination by foreground stars is very likely; the maximum-likelihood estimator, instead, has an almost flat error curve. Hence, our novel method is able to “recognize” the presence of foreground stars; moreover, the inclusion of the background density in the likelihood expression allows this estimator to “switch” to the number count technique in regions with large extinction. Figure 12, in particular, shows that even in extreme cases with a large foreground star contamination the maximum-likelihood method is still very reliable and accurate. To better appreciate this point, we note that, in our simulations, for $A_V = 25$ and $f = 0.2$ on average only one tenth of the ~ 25 stars are background to the cloud.

4.4. Limitations

The maximum-likelihood approach to the extinction measurements presents clear advantages with respect to the standard techniques in the simplified framework considered in this section (uniform A_V over the cloud patch analysed, thin-cloud approximation, simple model for the star intrinsic magnitude distribution). One could thus legitimately ask whether the maximum-likelihood technique is applicable to more realistic and complex situations.

4.4.1. Small-scale inhomogeneities

Most clouds present clear sign of substructure at different scales and a statistical analysis of the radio and NIR observations seems to indicate that these objects can be well described in terms of turbulent models (see, e.g.

Miesch & Bally 1994; Padoan et al. 1997). In presence of significant inhomogeneities, most methods (including the maximum-likelihood one) are expected to be biased toward low values because the background stars will no longer be randomly distributed in the patch of the sky used to estimate the local extinction value, but will be preferentially detected in low-extinction regions (cf. Eq. (43)). Although a detailed discussion of this effect is behind the scope of this paper, it is worth considering the following points:

- For the *HK* NICE, for the *JHK* NICER, and for the *JHK* maximum-likelihood methods, small scale inhomogeneities become important when the local variations of A_V on the patch of the sky considered are of order $\Delta = 1/(\alpha k_2 \ln 10)$, where $k_2 = A_H/A_V$ (cf. Eq. (8)). For typical 2MASS observations we find $\Delta \approx 7.3$ mag, and hence all methods should still be applicable to the analysis of regions with relatively low column densities (approximately $A_V < 10$ mag).
- The effect of substructures is studied in detail in a forthcoming paper (Lombardi 2005), where it is also presented a method to avoid the bias introduced by small-scale inhomogeneities. The application of this novel technique to a NICER map of the Pipe nebula confirms the expectations summarized in the last item. In particular, for the Pipe nebula the “standard” NICER has a negligible small bias (below 0.2 mag) for $A_V < 10$ mag, while the bias increases dramatically for larger extinction values (e.g., it reaches 1 mag at $A_V = 15$ mag). A similar bias behavior is expected for the maximum-likelihood technique described in this section.
- Since small-scale inhomogeneities are believed to be due to turbulent motions (Larson 1981; Padoan et al. 1997; Heyer & Brunt 2004), the probability distribution $p_{A_V}(A_V)$ is expected to be a log-normal:

$$p_{A_V}(A_V) = \frac{1}{\sqrt{2\pi}S A_V} \exp\left[-\frac{(\ln A_V - M)^2}{2S^2}\right]. \quad (65)$$

Independently from the exact form of $p_{A_V}(A_V)$, the maximum-likelihood approach can be used in this more general framework to obtain the relevant parameters of p_{A_V} (e.g., M and S in the case of the log-normal of Eq. (65)); moreover, the parameters estimated are asymptotically unbiased (see Eadie et al. 1971). We will consider the use of a non-trivial extinction probability distributions similar to the one of Eq. (65) in a follow-up paper.

4.4.2. Star magnitude distribution

In the implementation of the maximum-likelihood technique discussed in this section, we used the simple model for the magnitude probability distribution $p_M(\mathbf{M})$ (cf. Eq. (34)). However, the functional form of $p_M(\mathbf{M})$ used can be inaccurate in describing real data for several reasons:

- Different stellar populations can produce multiple peaks in the color distribution of stars. For example, giant and dwarf stars produce two distinct peaks in the $J - H$ histogram (the two peaks are also clearly visible in a *JHK* color–color diagram).

- Different stellar populations can also have different slopes of the number counts, which could differ significantly from the “nominal” value $\alpha \approx 0.34$ (see, e.g. Cambrésy et al. 2002). If the various stellar populations also have different intrinsic mean colors, then the simple model (34) could lead to inaccurate extinction measurements.
- The average NIR colors of stars in regions free of extinction are not completely independent of the star luminosity. For example, a color–magnitude plot of 2MASS stars shows that the average $J - K$ color slightly increases as the K magnitude increases. Since this effect is rather small, the associated bias is probably negligible in most cases; moreover, this effect can be included in the expression (35) with a suitable choice of the coefficients P , Q , and R .

All issues described above are strictly related to the simplified description for the probability $p_M(\mathbf{M})$ used here, and *not* to the maximum-likelihood method itself. In other words, it is possible (and relatively easy) to implement a maximum-likelihood estimator based on a more realistic probability distribution for the star magnitudes (e.g., synthetic stellar population models; see Robin et al. 2004; Jarrett et al. 1994). For this purpose, we note that the most computationally effective way to generalize $p_M(\mathbf{M})$ is by writing it as a linear combination of functions of the form (34) (with each function representing, de facto, a different star population).

4.4.3. General remarks

One of the most significant drawback of the maximum-likelihood approach is its speed. The implementation used in this paper is approximately one order of magnitude slower than the NICER method (at least on a typical workstation), and this might prevent large applications of the method proposed here. On the other hand, the fast technological progress in the computer speed justifies the work presented in this paper, in the sense that soon it will be possible to use the maximum-likelihood method on large datasets composed of millions of stars.

Another possible issue related to the technique presented here is the need for a more detailed knowledge of the properties of the data used. As a comparison, we note that the original NICE technique makes use only of the H and K magnitudes of stars and of the average color $\langle H - K \rangle$ of stars in the control field. In addition to these data, NICER also requires the estimated errors of star magnitudes in the NIR bands used. Finally, the maximum-likelihood method requires a detailed knowledge of the probability distribution $p_M(\mathbf{M})$, of the measurement errors $p_{\epsilon_i}(\hat{m}_i|m_i)$ of each star, and of the completeness functions $\hat{c}_\lambda(\hat{m}_\lambda)$. In reality, in the Bayesian approach implicitly adopted in this paper, the maximum-likelihood method can also be used if an *approximate* knowledge of these parameters is available. Suppose, for example, that the parameters P , Q , and R that characterize $p_M(\mathbf{M})$ (see Eq. (35)) are only approximately known from the measurements in the control field. In this case we can take these parameters as *unknowns* in the expression for the likelihood, and we can thus minimize L with respect to them as well as with respect to the parameters that characterize $p_{A_V}(A_V)$ (A_V and f in the case considered here).

As customary in standard maximum-likelihood problems, we include the knowledge on P , Q , and R as a *prior* in the function L . Note that this way the dimension of the space over which we need to minimize L greatly increases (and this can pose severe computational problems), but in principle the schema proposed is applicable to real cases.

Interestingly, the NICE and NICER technique can be seen as special cases of the maximum-likelihood method when no prior knowledge on R is available: the complete lack information on the normalization of $p_M(\mathbf{M})$ forces our method to use the only color information of the stars. Similarly, the number counts method can be recovered as special case when the knowledge on the average colors χ_a is absent (or, equivalently, when the scatter in the colors C_{ab} is very large; cf. Eq. (34)). This suggests that the NICE(R) techniques are not more robust of the maximum-likelihood one, but just simpler.

Finally, we mention that the presence of young stellar objects (that could be for example embedded in the dense cores) can in principle affect the extinction measurements. Hence, these objects should be “manually” removed before using any extinction measurement method, including the maximum-likelihood described in this paper.

5. Discussion and conclusions

In this paper we have considered the problem of an accurate determination of the extinction toward a dark molecular cloud. The results obtained here can be summarized in the following items:

- The extinction and the reddening of background stars have been considered from a general statistical point of view.
- The bias and uncertainties of the two main NIR techniques used to map the extinction, the star count and the color excess methods, have been discussed in detail. We have shown that, although the color excess method has generally a smaller error, it is affected by a large bias in presence of contamination by foreground stars. We have also shown that both NICE and NICER are affected by a relatively small bias (approximately 0.2 mag in A_V) as a result of selection effects.
- A new optimal maximum-likelihood method has been presented and tested with extensive simulations.

The simulations described in Sect. 4.3 have shown that *in the simple case of a thin cloud the maximum-likelihood estimator performs significantly better than the NICE and NICER estimators* (since the number count method is known to have a larger noise, we did not report a detailed comparison with this method here). However, the maximum-likelihood approach also allows us to consider more general cloud configurations, which cannot be easily dealt with using standard techniques. For example, the maximum-likelihood techniques could be used to measure directly on the same patch of the sky both the column density A_V and the fraction of foreground stars f ; alternatively, it would be also possible to determine f globally on the cloud and take it as a constant on the whole field (a good approximation for small clouds).

In Sect. 4.4.1 we discussed one of the most serious limitations of NIR color excess studies in molecular clouds, namely the bias introduced by substructures. In our original formulation, the maximum-likelihood method can be used not only in the thin cloud approximation, but in general for any functional form of $p_{A_V}(A_V)$. Hence, as mentioned above, we could implement the maximum-likelihood method using a more realistic probability distribution for A_V , such as the one of Eq. (65).

Another possibility offered by the maximum-likelihood approach is the generalization of the thin-cloud approximation to a multi-layer case, where two or more (thin) clouds located at different distances are observed on overlapping areas of the sky. For example, in case of a double cloud we could write (cf. Eq. (17))

$$p_{A_V}(A_V) = f^{(1)}\delta(A_V) + f^{(2)}\delta(A_V - A_V^{(1)}) + (1 - f^{(1)} - f^{(2)})\delta(A_V - A_V^{(1)} - A_V^{(2)}). \quad (66)$$

This configuration, for example, might be appropriate to study clouds close to the galactic center, where the superposition of different complexes is very likely. Such a method could effectively disentangle the effects of the two clouds provided the values of $f^{(1)}$ and $f^{(2)}$ are sufficiently different. Alternatively, one could use a double cloud as a null test, i.e. to check that indeed the thin-cloud approximation is appropriate (in this case one expects to find $A_V^{(1)} \simeq 0$, $A_V^{(2)} \simeq 0$, or $f^{(1)} \simeq 0$). Some of these possibilities will be investigated in detail in a follow-up paper by using 2MASS data.

Acknowledgements. We are very grateful to G. Bertin and J. Alves for useful and stimulating discussions, and to the referee, L. Cambresy for providing stimulating comments and for helping us improve the paper significantly.

References

- Alves, J., Lada, C. J., & Lada, E. A. 1999, ApJ, 515, 265
 Alves, J., Lada, C. J., & Lada, E. A. 2001, Nature, 409, 159
 Andreazza, C. M., & Vilas-Boas, J. W. S. 1996, A&AS, 116, 21
 Barnard, E. E. 1919, ApJ, 49, 1
 Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
 Bok, B. J. 1937, The distribution of the stars in space (Chicago: University of Chicago Press)
 Bok, B. J. 1956, AJ, 61, 309
 Cambr esy, L., Beichman, C. A., Jarrett, T. H., & Cutri, R. M. 2002, AJ, 123, 2559
 Cambr esy, L. 1999, A&A, 345, 965
 Cambresy, L., Epchtein, N., Copet, E., et al. 1997, A&A, 324, L5
 Casali, M. M. 1986, MNRAS, 223, 341
 Cressie, N. A. C. 1993, Statistics for Spatial Data (New York: Wiley & sons)
 Dickman, R. L. 1978, AJ, 83, 363
 Eadie, W., Drijard, D., James, F., Roos, M., & Sadoulet, B. 1971, Statistical Methods in Experimental Physics (Amsterdam, New-York, Oxford: North-Holland Publishing Company)
 Epchtein, N., de Batz, B., Capoani, L., et al. 1997, The Messenger, 87, 27
 Frerking, M. A., Langer, W. D., & Wilson, R. W. 1982, ApJ, 262, 590
 Harjunp a, P., Lehtinen, K., & Haikala, L. K. 2004, A&A, 421, 1087

- Heyer, M. H., & Brunt, C. M. 2004, *ApJ*, 615, L45
- Horner, D. J., Lada, E. A., & Lada, C. J. 1997, *AJ*, 113, 1788
- Indebetouw, R., Mathis, J. S., Babler, B. L., et al. 2004, *ArXiv Astrophysics e-prints*
- Jarrett, T. H., Dickman, R. L., & Herbst, W. 1994, *ApJ*, 424, 852
- Jones, T. J., Hyland, A. R., & Bailey, J. 1984, *ApJ*, 282, 675
- Jones, T. J., Hyland, A. R., Robinson, G., Smith, R., & Thomas, J. 1980, *ApJ*, 242, 132
- Kleinmann, S. G., Lysaght, M. G., Pughe, W. L., et al. 1994, *Exp. Astron.*, 3, 65
- Lada, C. J., Alves, J., & Lada, E. A. 1999, *ApJ*, 512, 250
- Lada, C. J., Lada, E. A., Clemens, D. P., & Bally, J. 1994, *ApJ*, 429, 694
- Larson, R. B. 1981, *MNRAS*, 194, 809
- Li, A., & Greenberg, J. M. 2003, in *Solid State Astrochemistry*, 37, [arXiv:astro-ph/0204392]
- Lilley, A. E. 1955, *ApJ*, 121, 559
- Lombardi, M. 2005, *A&A*, submitted
- Lombardi, M., & Alves, J. 2001, *A&A*, 377, 1023
- Mathis, J. S. 1990, *ARA&A*, 28, 37
- Mattila, K. 1986, *A&A*, 160, 157
- Miesch, M. S., & Bally, J. 1994, *ApJ*, 429, 645
- Padoan, P., Jones, B. J. T., & Nordlund, A. P. 1997, *ApJ*, 474, 730
- Rieke, G. H., & Lebofsky, M. J. 1985, *ApJ*, 288, 618
- Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2004, *A&A*, 416, 157
- Schultheis, M., Ganesh, S., Simon, G., et al. 1999, *A&A*, 349, L69
- Trumpler, R. J. 1930, *Lick Observatory Bulletin*, 14, 154
- van de Hulst, H. C. 1946, *Recherches Astronomiques de l'Observatoire d'Utrecht*, 11, 2
- Wolf, M. 1923, *Astron. Nachr.*, 219, 109

Online Material

Appendix A: The median and related estimators

The goal of this appendix is to derive the probability distribution of the median of n independent identical random variables. This analysis is useful to address some of the issues discussed in Sect. 3.2.

A.1. Notation

In the following we will often deal with *ordered* and *unordered* n -tuples. The latter can be taken to be a set of n elements, where typically each element is a random variable or an element of another tuple; the former can be taken to be a list of n elements, where thus each element is associated to a position in the list. Hence, given a positive integer $k \leq n$, it makes sense to talk about the k th element of an ordered n -tuple, while this is meaningless for an unordered tuple.

We will denote an ordered n -tuple using brackets, as in $[x_1, \dots, x_n]$; instead, we will reserve braces for the unordered tuples, as in $\{x_1, \dots, x_n\}$. Note that, for consistency with the definition, we identify unordered tuples if these differ just by a permutation of the elements: thus $\{x_1, x_2, x_3\}$ is, for instance, identical to $\{x_2, x_1, x_3\}$.

A.2. P_k^n and the median estimator

Let us call $p(x)$ the probability distribution for the real random variable x , and let us consider the generation of unordered n -tuples $\{x_1, x_2, \dots, x_n\}$ of such variables. Suppose that, after the random generation of a tuple, we order the tuple such that $x_1 \leq x_2 \leq \dots \leq x_n$. We consider now the probability distribution $p_k^n(x_k)$ for the k th element of the ordered tuple $[x_1, x_2, \dots, x_n]$, which can be written as

$$p_k^n(x_k) = n \binom{n-1}{k-1} p(x_k) [P(x_k)]^{k-1} [1 - P(x_k)]^{n-k}, \quad (\text{A.1})$$

where $P(x)$ is the cumulative probability distribution of x :

$$P(x) = \int_{-\infty}^x p(x') dx'. \quad (\text{A.2})$$

The expression appearing in Eq. (A.1) can be explained as follows. Let us consider an element (for example the first) of the unordered n -tuple. The probability that this element is the k th in the ordered tuple and that it has a value in the range $[x, x+dx]$ is the product of three terms:

- $p(x) dx$ to takes into account the intrinsic probability distribution of x ;
- $\binom{n-1}{k-1} [P(x_k)]^{k-1} [1 - P(x_k)]^{n-k}$, which is a simple binomial distribution giving the probability that the value chosen has $k-1$ elements of the order n -tuple at its left, and $n-k$ elements at its right;
- finally, we have to multiply the whole result by n in order to take into account the arbitrary choice of the k th element in the n -tuple.

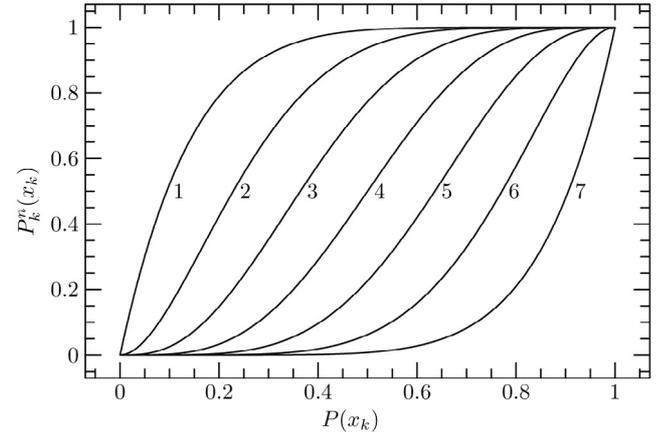


Fig. A.1. The cumulative probability distribution $P_k^n(x_k)$ as a function of $P(x)$ for $n = 7$ and various values of k (marked close to the relative curves).

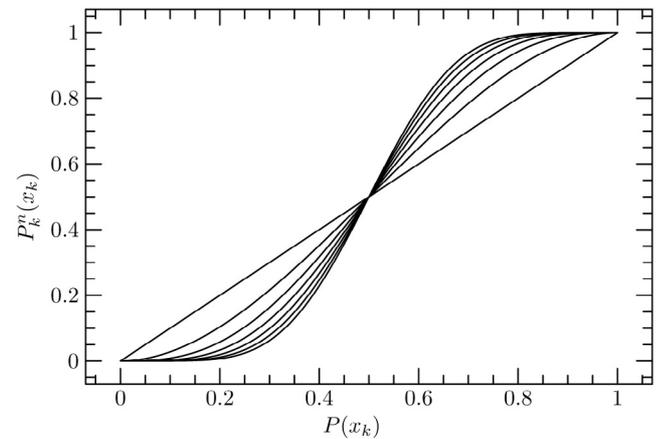


Fig. A.2. The cumulative probability distribution $P_k^n(x_k)$ as a function of $P(x)$ for $k \in \{1, 2, 3, 4, 5, 6, 7\}$ and $n = 2k - 1$. Note that all curves pass through the point $(0.5, 0.5)$.

The cumulative distribution $P_k^n(x_k)$ is given by

$$\begin{aligned} P_k^n(x_k) &= \int_{-\infty}^{x_k} p_k^n(x'_k) dx'_k \\ &= \frac{n!}{(k-1)!(n-k)!} \sum_{\ell=0}^{n-k} (-1)^\ell \binom{n-k}{\ell} \\ &\quad \times \int_{-\infty}^{x_k} [P(x'_k)]^{k-1+\ell} p(x'_k) dx'_k \\ &= \sum_{m=k}^n (-1)^{m+k} \binom{n}{m} \binom{m-1}{k-1} [P(x_k)]^m. \end{aligned} \quad (\text{A.3})$$

In the last step we changed the index variable into $m = k + \ell$. The final result obtained in Eq. (A.3) has the advantage to be a simple (polynomial) expression in $P(x)$.

Figures A.1 and A.2 show the polynomials $P_k^n(x_k)$ as a function of $P(x_k)$ for various values of n and k . These figures suggest a number of properties for $P_k^n(x_k)$ that can be verified analytically with the help of the equations written above:

- $P_k^n(x) = 0$ if $P(x) = 0$, and $P_k^n(x) = 1$ if $P(x) = 1$; moreover, P_k^n is a monotonic increasing function of P . This implies that $p_k^n(x)$ vanishes where $p(x)$ vanishes.

- Using Eq. (A.1), it is possible to show that (see below Eq. (A.10))

$$\frac{1}{n} \sum_{k=1}^n P_k^n(x) = P(x), \quad \frac{1}{n} \sum_{k=1}^n p_k^n(x) = p(x). \quad (\text{A.4})$$

- As suggested by Fig. A.1 and by Eq. (A.1), we have

$$P_k^n(x) = 1 - P_{n+1-k}^n(x'), \quad (\text{A.5})$$

provided that $P(x) = 1 - P(x')$. This, in particular, implies that $P_k^{2k-1}(x) = 1/2$ if $P(x) = 1/2$ (see Fig. A.2).

- Equation (A.3) specialized to the cases $k = 1$ and $k = n$ is

$$P_1^n(x_1) = [P(x_1)]^n, \quad P_n^n(x_n) = 1 - [1 - P(x_n)]^n. \quad (\text{A.6})$$

The two last properties have a close relationship with the median estimator. We note, in fact, that for odd n the median estimator is in our notation $x_{(n+1)/2}$; as a result, the probability distribution $p_k^{2k-1}(x_k)$ is just the probability distribution of the median for $n = 2k - 1$. Hence, the property (A.5) written above can be rephrased as

The median of the probability distribution of the median estimator of the random variable x is the median of the probability distribution of x .

Note that here we use the term ‘‘median’’ to denote both the usual median of an n -tuple $\{x_1, \dots, x_n\}$, and the median of a distribution, defined as the value x such that the value cumulative distribution is $1/2$.

A.3. $P_{\leq k}^n$ and $P_{\geq k}^n$

For our purposes, it is also useful to define and study two probability distributions closely related to $P_k^n(x_k)$. Let us consider again the ordered n -tuple $[x_1, \dots, x_n]$, where each element is drawn from a probability distribution $p(x)$. Now let us retain only the first k elements of the ordered n -tuple, and let us call $p_{\leq k}^n(x_{\leq k})$ the probability distribution for each element of the *unordered* k -tuple $\{x_1, \dots, x_k\}$; similarly, we call $p_{\geq k}^n(x_{\geq k})$ the probability distribution for the tuple $\{x_k, \dots, x_n\}$ where only the last $(n - k + 1)$ elements of the original ordered n -tuple are retained.

The evaluation of the probability distribution $p_{\geq k}^n(x)$ can be broken into two parts. Consider the $(k - 1)$ th element x_{k-1} of the tuple $[x_1, \dots, x_n]$ (i.e., the last element discarded); then, clearly, each element of the $(n - k + 1)$ -tuple $\{x_k, \dots, x_n\}$ cannot be smaller than x_{k-1} . Moreover, each element of this *unordered* tuple is distributed between x_{k-1} and ∞ according to the (truncated) original probability distribution $p(x)$. Finally, by repeating this argument for each possible value of x_{k-1} (weighted by p_{k-1}^n), we obtain the expression

$$P_{\geq k}^n(x) = \int_{-\infty}^x p_{k-1}^n(x_{k-1}) \frac{p(x)}{1 - P(x_{k-1})} dx_{k-1}. \quad (\text{A.7})$$

Note that the term $1 - P(x_{k-1})$ is introduced here to correctly normalize the truncated probability distribution $p(x)$.

As usual in this appendix, it is more convenient to consider the cumulative probability distributions. We can thus integrate

$p_{\geq k}^n(x)$ over x and obtain, after a few manipulations, a closed expression for $P_{\geq k}^n(x)$. Here, however, we prefer to follow a different and simpler path.

Let us consider again the tuple $[x_k, \dots, x_n]$. Each element of this *ordered* tuple follows the probability distribution $p_k^n(x_k)$ discussed in the previous section. Hence, if we consider the *unordered* tuple, the elements $\{x_k, \dots, x_n\}$ will follow the *average* distribution

$$P_{\geq k}^n(x) = \frac{1}{n - k + 1} \sum_{k'=k}^n p_{k'}^n(x). \quad (\text{A.8})$$

By integrating both sides of this equation on dx , we can verify that the same relation holds for the cumulative probability distributions. Let us then evaluate the sum

$$\begin{aligned} P_{\geq k}^n(x) &= \frac{1}{n - k + 1} \sum_{k'=k}^n P_{k'}^n(x) \\ &= \frac{1}{n - k + 1} \sum_{m=k}^n \binom{n}{m} [P(x)]^m \\ &\quad \times \sum_{k'=k}^m (-1)^{k+m} \binom{m-1}{k'-1}. \end{aligned} \quad (\text{A.9})$$

We now consider two cases. If $k = 1$, then the last sum can be taken to be the binomial expansion of $(1 - 1)^{m-1}$, which vanishes for all integers $m > 1$. Hence we are left just with the case $m = 1$, for which we obtain

$$P_{\geq 1}^n = \frac{1}{n} \sum_{k'=k}^n P_{k'}^n(x) = P(x). \quad (\text{A.10})$$

This shows that each element of the *unordered* n -tuple $\{x_1, \dots, x_n\}$ follows the original distribution $p(x)$, a very natural result indeed.

If, instead, $k > 1$, then using the properties of the binomial coefficient we find

$$P_{\geq k}^n(x) = \frac{1}{n - k + 1} \sum_{m=k}^n (-1)^{m+k} \binom{n}{m} \binom{m-2}{k-2} [P(x)]^m. \quad (\text{A.11})$$

The similarity of this result with the last line of Eq. (A.3) is rather surprising.

Similarly, we wish to investigate the cumulative probability distribution $P_{\leq k}^n(x)$ for the k -tuple $\{x_1, \dots, x_k\}$. This quantity is better evaluated using $P_{\geq k}^n(x)$:

$$\begin{aligned} P_{\leq k}^n(x) &= \frac{1}{k} \sum_{k'=1}^k P_{k'}^n(x) \\ &= \frac{1}{k} \left[\sum_{k'=1}^n P_{k'}^n(x) - \sum_{k'=k+1}^n P_{k'}^n(x) \right] \\ &= \frac{1}{k} \left[nP(x) - (n - k)P_{\geq k+1}^n(x) \right] \\ &= \frac{n}{k} P(x) + \frac{1}{k} \sum_{m=k+1}^n \binom{n}{m} \binom{m-2}{k-1} (-1)^{m+k} \\ &\quad \times [P(x)]^m. \end{aligned} \quad (\text{A.12})$$

As an application of the results obtained in this section, we evaluate the bias introduced in the estimate of the column density when using the technique described in Sect. 3.2 to remove foreground stars (see Eq. (54)). Suppose that we observe stars in a region with no significant extinction, so that both foreground and background stars have the same distribution in colors. For simplicity, we assume that the reddening estimates \hat{A}_V for each individual star is a Gaussian distribution with vanishing average and variance $\sigma_{\hat{A}_V}$. In this situation, if we exclude the k bluer stars (because they are taken to be foreground), we will bias the estimate of A_V toward large column densities. In particular, the column densities for the stars left will be distributed according to $p_{\geq k+1}^n$. The median of this distribution can be evaluated as follows. At $A_V = 0$ we have $P(A_V) = 1/2$, while

$$P_{\geq k+1}^n(0) = \frac{1}{n-k} \left[\frac{n}{2} - \sum_{k'=1}^k P_{k'}^n(1/2) \right]. \quad (\text{A.13})$$

Assuming $n \gg k$, then each term in the sum above is approximately unity (cf. above Fig. A.1). Hence we obtain

$$P_{\geq k+1}^n(0) \simeq \frac{n-2k}{2n-2k}. \quad (\text{A.14})$$

The fact that this quantity is not exactly 1/2 is telling us that there is a bias. The exact amount of this bias can be evaluated by solving the equation $P_{\geq k+1}^n(x) = 1/2$. Using Newton's method we obtain then the approximate solution

$$x \simeq \left[\frac{dP_{\geq k+1}^n}{dx} \right]_{x=0}^{-1} \left[\frac{1}{2} - P_{\geq k+1}^n(0) \right] \simeq \frac{k}{2np(0)}. \quad (\text{A.15})$$

Note that in this equation we used the relationship between cumulative and differential probability distributions; moreover, we simplified $p_k^n(0) \simeq 0$ at low k , and retained only terms linear in k/n . The final result obtained is thus given by Eq. (54).