On the use of conditional expectation estimators

Sergio Ortobelli, Tommaso Lando

Abstract— This paper discusses two different methods to estimate the conditional expectation. The kernel non-parametric regression method allows to estimate the regression function, which is a realization of the conditional expectation \( E(Y|X) \). A recent alternative approach consists in estimating the conditional expectation (intended as a random variable), based on an appropriate approximation of the \( \sigma \)-algebra generated by \( X \). In this paper, we propose a new procedure to estimate the distribution of the conditional expectation based on the kernel method, so that it is possible to compare the two approaches by verifying which one better estimates the true distribution of \( E(Y|X) \). In particular, if we assume that the two-dimensional variable \((X, Y)\) is normally distributed, then the true distribution of \( E(Y|X) \) can be computed quite easily, and the comparison can be performed in terms of goodness-of-fit tests.

Keywords—Conditional Expectation, Kernel, Non Parametric, Regression.

I. INTRODUCTION

Within a bivariate probabilistic framework, this paper discusses different methods to estimate the conditional expected value. On the one hand, several well known methods are aimed at estimating the regression function \( g(x) = E(Y|X = x) \), which represents a realization of the random variable \( E(Y|X) \). In particular, the kernel non-parametric regression (see [1] and [2]) allows to estimate \( E(Y|X = x) \) as a locally weighted average, based on the choice of an appropriate kernel function: the method yields consistent estimators, provided that the kernel functions and the random variable \( Y \) satisfy some conditions, described in Section II. On the other hand, an alternative methodology was recently introduced by [3] for estimating the random variable \( E(Y|X) \): this method has been proved to be consistent without requiring any regularity assumption. In this paper we stress the difference between the two methods, that are actually aimed at different estimates (i.e. the mathematical function \( g(x) \) vs. the random variable \( E(Y|X) \)) and therefore are not comparable. In order to compare these two different methodologies, we propose a method to estimate the distribution of \( E(Y|X) \) based on the kernel non-parametric formula proposed by [1] and [2]. Then, if we know the real distribution of \( E(Y|X) \) (which, for instance, can be easily computed in case of normality), then we can perform a simulation analysis, drawing a bivariate random sample from \((X, Y)\), and finally investigate which estimated distribution better fits to the true one.

The paper is organized as follows: in Section II we present the different methodologies and their properties; in Section III we examine a method to compare the two estimators, with assumption of normality; in Section IV we briefly illustrate the financial interpretation and possible application of the conditional expected value.

II. METHODS

In this section we describe two different procedures to evaluate the conditional expected value between two random variables. Let \( X : \Omega \rightarrow \mathbb{R} \) and \( Y : \Omega \rightarrow \mathbb{R} \) be integrable random variables in the probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Let \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\) be a random sample of independent observations from the bi-dimensional variable \((X, Y)\). The first procedure is aimed at estimating the conditional expectation of \( Y \) given \( X = x \), which is a mathematical function of \( X \); the second method yields an unbiased and consistent estimator of the random variable \( E(Y|X) \).

The kernel non-parametric regression

It is well known that, if we know the form of the function \( g(x) = E(Y|X = x) \) (e.g. polynomial, exponential, etc.), then we can estimate the unknown parameters of \( g(x) \) with several methods (e.g. least squares). In particular, if we do not know the general form of \( g(x) \), except that it is a continuous and smooth function, then we can approximate it with a non-parametric method, as proposed by [1] and [2]. Thus, \( g(x) \) can be estimated by:

\[
\hat{g}_n(x) = \frac{\sum_{i=1}^{n} y_i K(x-x_i/h_n)}{\sum_{i=1}^{n} K(x-x_i/h_n)},
\]

where \( K(x) \) is a density function such that i) \( K(x) < C < \infty \); ii) \( \lim_{x \to \pm \infty} |xK(x)| = 0 \); iii) \( h(n) \to 0 \) when \( n \to \infty \). The function \( K(x) \) is denoted by kernel, observe that kernel functions are generally used for estimating probability densities non-parametrically (see [4]). It was proved in [1] that if \( Y \) is quadratically integrable then \( \hat{g}_n(x) \) is a consistent estimator for \( g(x) \). In particular, observe that, if we denote by
The joint density of $(X, Y)$, the denominator of (1) converges to the marginal density of $X \int f(x, y)dy$, while the numerator converges to the function $\int yf(x, y)dy = \int_\infty^\infty \int_{\{x=x\}} yP(dx, dy)$ (note that, if $X$ is continuous, the function $\int_{\{x=x\}} yP(dx, dy)/\int_\infty^\infty \int_{\{x=x\}} P(dx, dy)$ has to be intended as a regular conditional probability).

The OLP method

We now describe an alternative non-parametric approach [3] for approximating the conditional expectation, the method is denoted by “OLP”, which is an acronym of the authors’ names. Define by $\mathfrak{F}_X$ the $\sigma$-algebra generated by $X$ (that is, $\mathfrak{F}_X = \sigma(X) = X^{-1}(B) = \{X^{-1}(B) : B \in B\}$, where $B$ is the Borel $\sigma$-algebra on $\mathbb{R}$). Observe that the regression function is just a “pointwise” realization of the random variable $E(Y|\mathfrak{F}_X)$, which can equivalently be denoted by $E(Y|X)$. The following methodology is aimed at estimating $E(Y|X)$ rather than $g(x)$.

$\mathfrak{F}_X$ can be approximated by a $\sigma$-algebra generated by a suitable partition of $\Omega$. In particular, for any $k \in \mathbb{N}$, we consider the partition $\{A_j\}_{j=1}^k = \{A_1, ..., A_k\}$ of $\Omega$ in $b^k$ subsets, where $b$ is an integer number greater than 1 and:

- $A_1 = \{\omega: X(\omega) \leq F_\mu^{-1}\left(\frac{1}{b}\right)\}$,
- $A_h = \{\omega: F_\mu^{-1}\left(\frac{h}{b}\right) < X(\omega) \leq F_\mu^{-1}\left(\frac{h+1}{b}\right)\}$, for $h = 2, ..., b^k - 1$,
- $A_{b^k} = \Omega - \cup_{h=1}^{b^k} A_h = \{\omega: X(\omega) > F_\mu^{-1}\left(\frac{b^k}{b}\right)\}$.

Starting with the trivial sigma algebra $\mathfrak{F}_0 = \{\emptyset, \Omega\}$, we can obtain a sequence of sigma algebras generated by these partitions, for different values of $k$ ($k=1, ..., m, ...$). For instance, $\mathfrak{F}_1 = \sigma(\emptyset, \Omega, A_1, ..., A_k)$ is the sigma algebra generated by $A_1 = \{\omega: X(\omega) \leq F_\mu^{-1}(1/b)\}$, $A_2 = \{\omega: F_\mu^{-1}\left(\frac{1}{b}\right) < X(\omega) \leq F_\mu^{-1}\left(\frac{2}{b}\right)\}$, $A_{b-1} = \{\omega: F_\mu^{-1}\left(\frac{b-1}{b}\right) < X(\omega) \leq F_\mu^{-1}(1)\}$, and $A_b = \{\omega: X(\omega) > F_\mu^{-1}(1)\}$. Generally:

$$\mathfrak{F}_k = \sigma \left( \bigcup_{j=1}^{b^k} A_j \right), k \in \mathbb{N}. \tag{2}$$

Hence, it is possible to estimate the random variable $E(Y|\mathfrak{F}_X)$ by

$$E(Y|\mathfrak{F}_X)(\omega) = \sum_{j=1}^{b^k} \frac{1_{A_j}(\omega)}{P(A_j)} \int_{A_j} YdP = \sum_{j=1}^{b^k} \frac{1}{P(A_j)} E(Y|A_j) 1_{A_j}(\omega), \tag{3}$$

where $1_{A_j}(\omega) = \begin{cases} 1, & \omega \in A_j \\ 0, & \omega \notin A_j \end{cases}$. Indeed, by definition of the conditional expectation, we can easily verify that $E(Y|\mathfrak{F}_X)$ is the unique $\mathfrak{F}_X$-measurable function such that, for any set $A \in \mathfrak{F}_X$, (that can be seen as a union of disjoint sets, in particular $A = \cup_{A_j \in \mathfrak{F}_X} A_j$) we obtain the equality

$$\int_A E(Y|\mathfrak{F}_X)dP = \int_A Y(\omega)dP(\omega) \tag{4}$$

It is proved in [3] that $E(Y|\mathfrak{F}_X)$ is a consistent estimator of the random variable $E(Y|X)$, that is, $\lim_{k \to \infty} E(Y|\mathfrak{F}_X) = E(Y|X)$ a.s.

The method, as defined by (3), requires only that $Y$ is an integrable random variable. From a practical point of view, given $n$ i.i.d. observations of $Y$, if we know the probability $p_i$ corresponding to the $i$-th outcome $Y_i$, we obtain:

$$E(Y|A_j) = \frac{\sum_{j \in A_j} Y_i p_i}{P(A_j)}. \tag{5}$$

Otherwise, we can give uniform weight to each observation, which yields the following consistent estimator of $E(Y|A_j)$:

$$\frac{1}{n_{A_j}} \sum_{j \in A_j} Y_i, \tag{6}$$

where $n_{A_j}$ is the number of elements of $A_j$. Therefore, we are always able to estimate $E(Y|\mathfrak{F}_X)$, which in turn is a consistent estimator of the conditional expected value $E(Y|X)$.

III. COMPARISON IN CASE OF NORMALITY

If we assume that $X$ and $Y$ are jointly normally distributed, i.e. $(X, Y) \sim N(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, we can obtain the distribution of the random variable $E(Y|X)$ quite easily. Indeed, we know that

$$g(x) = E(Y|X = x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X). \tag{7}$$

therefore, as $X \sim N(\mu_X, \sigma_X)$, we obtain that

$$E(Y|X) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (X - \mu_X) - N(\mu_Y, |\rho|\sigma_Y). \tag{8}$$

Of course, if we simulate data from $(X, Y)$ and approximate $E(Y|X)$ with the estimator $E(Y|\mathfrak{F}_X)$ defined in (3), we can finally compare the true and the theoretical (estimated) distribution by performing a goodness-of-fit test. Differently, the kernel non-parametric regression method does not allow to estimate $E(Y|X)$, but only yields a consistent estimator of $g(x)$. However, assume that the random variable $X'$ is independent from $X$ and moreover $X = \epsilon X'$ (that is, $X' \sim N(\mu_X, \sigma_X)$ and $\rho(X, X') = 0$); in this case we can estimate $E(Y|X')$ with

$$g_n(x') = \frac{\sum_{j=1}^{n_{A_j}} Y_j \phi\left(\frac{x' - \mu_X}{\sigma_X}\right)}{\sum_{j=1}^{n_{A_j}} \phi\left(\frac{x' - \mu_X}{\sigma_X}\right)}, \tag{9}$$

and thereby we can also estimate the distribution of $E(Y|X)$, because $E(Y|X') = \epsilon E(Y|X)$. Obviously, the estimate depends on the choice of the kernel function $K$. It is proved that $E(Y|\mathfrak{F}_X)$ converges almost surely to $E(Y|X)$ ($E(Y|\mathfrak{F}_X) \to_{a.s.} E(Y|X)$). Moreover, note that also $g_n(X')$ satisfies a weaker convergence property (convergence in distribution). Indeed, we have that

$$g_n(X') \to_{d.s.} E(Y|X') = \epsilon E(Y|X), \tag{10}$$

thus we obtain that $g_n(X') \to_{a.s.} E(Y|X)$.

Finally, it is possible to compare the two methods by verifying which one better estimates the distribution of $E(Y|X)$, future studies will be focused on this issue.

IV. CONDITIONAL EXPECTATION AND FINANCIAL APPLICATIONS

The conditional expectation of a random variable given another can be especially useful for financial applications. In particular, we can use conditional expectation estimators.
either for ordering the investors choices as suggested by [3] or to evaluate and exercise those arbitrage opportunities when applies in the market.

We recall the classical definitions of first and second-orders stochastic dominance.

First order stochastic dominance: \( X \) FSD \( Y \) if and only if \( F_X(t) \leq F_Y(t), \forall t \in \mathbb{R} \)
or, equivalently \( X \) FSD \( Y \) if and only if \( E(g(X)) \geq E(g(Y)) \)for any increasing function \( g \).

Second order stochastic dominance (increasing concave order): \( X \) SSD \( Y \) if and only if \( \int_{-\infty}^{\infty} F_X(u)du \leq \int_{-\infty}^{\infty} F_Y(u)du, \forall t \in \mathbb{R} \) or, equivalently \( X \) SSD \( Y \) if and only if \( E(g(X)) \geq E(g(Y)) \) for any increasing and concave function \( g \).

Obviously \( X \) FSD \( Y \) implies also \( X \) SSD \( Y \).

If we assume that \( X \) and \( Y \) are, for instance, two different gambles or investments, the financial interpretation of stochastic orders follows straightforward. Indeed, in this case \( X \) FSD \( Y \) means that \( X \) is stochastically “larger” than \( Y \), while \( X \) SSD \( Y \) indicates a larger expectation of gain and generally an inferior “risk”. Those investors who prefer \( X \) to \( Y \), provided that \( X \) SSD \( Y \), are generally defined non-satisfiable risk averse investors. The following property characterizes the second order stochastic dominance in terms of conditional expectation.

Super-martingale property. \( X \) SSD \( Y \) if and only if there exist two random variables \( X', Y' \) defined on the same probability space that have the same distribution of \( X \) and \( Y \) such that:

\[
E(Y'|X') \leq X' \text{ a.s.}.
\]

The proof of this property arises from the analysis proposed by Strassen [5] and is a well-known result of ordering theory (see also [6]-[7] and the references therein).

Thus, using the empirical evaluation of the conditional expected value, we can attempt to order the investors choices as suggested by [3]. On the other hand, using the fundamental theorem of arbitrage, we know that there exist no arbitrage opportunities in the market if there exists a risk neutral martingale measure under which the discounted price process results a martingale. So, when we assume that the filtration \( \{\mathcal{F}_t\} \) is the one generated by the price process \( \{X_t\}_t \) (assumed to be a Markov process) then we get that \( E(X_t|\mathcal{F}_t) = E(X_t|\mathcal{F}_0) \). Therefore, this property the conditional expected value estimator and the fundamental theorem of arbitrage can be used to estimate the risk neutral measure and the presence of arbitrage opportunities in the market.

V. CONCLUSION

In this paper, we deal with two methodologies for estimating the conditional expectation, studying their properties and analyzing their differences. We also propose a procedure to estimate the distribution of \( E(Y|X) \) based on the kernel method. We observe that the OLP method proposed by [3] yields a consistent estimator of the random variable \( E(Y|X) \), while the generalized kernel method, proposed in eq. (9) yields a consistent estimator of the distribution function of \( E(Y|X) \). In future work it will be possible to compare these two methodologies by verifying which one better estimates the distribution of \( E(Y|X) \), based on simulation analysis and goodness-of-fit tests. Moreover, we recall that these estimators may have several financial applications such as ordering investors’ opportunities or identifying arbitrage opportunities.

REFERENCES