Minimizing Risks When There Is No Faith in a Distribution Model

Erick Delage and Yinyu Ye

Whether we like it or not, our fortune depends heavily on the fluctuations of the financial markets. Indeed, most of our long-term plans rely on our having the financial means to achieve them. When preparing for their execution, we necessarily seek a risk-free investment that will at least follow inflation. Unfortunately, the multiple financial crises we have witnessed in the last 10 years have cast some doubt on the existence of such investments. Taking as an example the largest pension fund in Canada, that is, the Caisse de dépôt et placement du Québec, a simple analysis of recent yearly reports indicates that, for the contributions made to the fund in 2000, it made the equivalent of 4% in yearly returns (a mere 2% above yearly inflation rates for those same years). Yet, the value of this fund's portfolio has been almost as volatile as the market: it incurred losses of 25% in 2008 and of over 5% in both 2001 and 2002. In an attempt to rationalize such a generalized failure to harness the risk variable, many have placed the blame on blind faith in a distribution model, accusing, for example, the log-normal return model behind the pricing of options, or the Gaussian copula model behind CDOs.

This phenomenon doesn't occur only in finance; it can be observed more generally in many decision-making problems that involve uncertainty. Indeed, the most popular approach for dealing with such problems starts by defining a probability distribution to characterize the relative likelihood of each potential outcome. The problem is that we often don't have the resources to do this properly. In some cases, there just isn't enough of a budget to collect a large number of observations or to consult with different experts; while, in other cases, using a popular distribution model simply makes the problem much easier to solve. Whatever the reason, it is important to realize that when we choose to use a distribution model without a sound rationale, we must then bear the responsibility for placing our faith in that model. This means that it should not come as a surprise if the exposure to risk ends up being greater than expected.

In this paper, we study an approach that circumvents the need to commit to a probabilistic model when developing a decision model. Instead, this approach requires that the decision-maker identify the known characteristics of the distribution. For instance, one could start by identifying regions in which all outcomes, the mean and the covariance matrix of the random parameters should fall. These characteristics are then used to define a set of potential distributions (whose size will depend on the amount of information available). The risks related to a decision are then measured with respect to the worst choice of distributions in this set. In particular, we show that, for a wide range of problems, if the information is limited to the support, mean, and covariance matrix, then the optimal decision can be found very efficiently. This is surprising, since for many of these cases, solving the problem for a single distribution is already a difficult computational task.

We also propose a methodology that can be used to estimate the characteristics of a distribution based on historical data. Given that each sample is independent and identically distributed, and that the magnitude of the vector of random parameters cannot be too large, the "distributional set" that is derived is guaranteed with a high probability to contain the true distribution behind the historical samples. One therefore has probabilistic guarantees that the measured risk is a good approximation of the true risk, and that risk is not underestimated. In effect, these results give us the tools we need to renounce having to place our faith in a distribution model and instead to rely on the well-established theory of probability.

> In an attempt to rationalize such a generalized failure to harness the risk variable, many have placed the blame on blind faith in a distribution model, accusing, for example, the log-normal return model behind the pricing of options, or the Gaussian copula model behind CDOs.

This "distributionally robust methodology" was evaluated on a portfolio-selection problem with historical data downloaded from *Yahoo! Finance*; where 30 assets were followed over the years 2001 to 2007. The results provided convincing evidence that while approaches relying on a distribution model (in particular, the empirical distribution model) can achieve positive returns, exposure to risk can be significantly reduced without any loss in return simply by recognizing that distributions are fallible. ■

Operations Research, 58(3), pp. 596-612, 2010.
Original title
Distributionally Robust Optimization under Moment Uncertainty with Application to Data-Driven Problems

Erick Delage
Department of Management Sciences
HEC Montréal and GERAD

Yinyu Ye
Department of Management Science and Engineering
Stanford University, California

Comment minimiser le risque sans se fier à une distribution

Erick Delage et Yinyu Ye

Que ça nous plaise ou non, notre fortune dépend fortement des fluctuations des marchés financiers. En effet, le succès d'une grande part de nos plans à long terme repose sur la présence de moyens financiers suffisants. Alors que nous tentons d'accumuler ces fonds, il est important de pouvoir investir ses économies sans risque afin qu'elles suivent au moins le taux d'inflation. Malheureusement, les événements dont nous avons été témoins au cours des dernières années ont semé le doute sur l'existence même d'un tel type d'investissement. Prenons comme exemple le cas du plus important fonds de pension canadien, i.e. la Caisse de dépôt et placement du Québec. Une analyse rapide de ses rapports de rendements annuels nous indique que pour les placements faits en 2000, le fonds a atteint l'équivalent d'un rendement annuel de 4 % (un maigre 2 % au-dessus du taux d'inflation annuel). Ce chiffre est d'autant plus inquiétant que, durant ces années, les rendements de ce fonds furent presque aussi volatiles que le marché, accumulant des pertes de 25 % en 2008, et de plus de 5 % en 2001 et en 2002. Afin d'expliquer la surexposition au risque de plusieurs fonds de ce type, nombreux sont ceux qui ont attribué la situation à la confiance aveugle des gestionnaires de portefeuilles en leur modèle stochastique –tenant responsable, par exemple, l'utilisation du modèle log-normal en faveur de la tarification des options ou du modèle de copule gaussienne en faveur des CDO (*Collateralized Debt Obligations* : obligations adossées à des actifs).

Cette remise en question vis-à-vis du choix d'un modèle probabiliste ne devrait pas être limitée au monde de la finance. En fait, la question se pose pour toute situation de prise de décision en contexte d'incertitude. En effet, le choix d'un modèle stochastique constitue la première étape de plusieurs méthodes populaires. Le problème réside dans le fait que nous n'avons pas toujours les ressources pour mener cette étape proprement à terme. Dans certains cas, ce sont les ressources financières qui nous limitent quant à la qualité des experts que nous consultons ou quant à la quantité de données que nous collectons. Dans d'autres cas, avoir recours au modèle stochastique « à la mode » rend tout simplement le problème plus facile à résoudre. Peu importe la raison, il est important de reconnaître que si l'on est incapable de justifier son choix de modèle stochastique, il faut nécessairement assumer la responsabilité de l'avoir choisi. Ceci signifie qu'il ne faut pas s'étonner si notre exposition au risque s'avère plus élevée que prévu.

Dans cette communication, nous étudions une approche qui permet de se soustraire au besoin d'adopter un modèle stochastique précis lors du développement d'un modèle décisionnel. L'approche propose plutôt d'identifier certaines caractéristiques de la distribution. On peut, par exemple, décrire des régions dans lesquelles on pense retrouver les valeurs incertaines, leurs moyennes ainsi que leurs variances et covariances. Ces caractéristiques sont ensuite utilisées pour définir un ensemble de distributions plausibles (dont la taille dépendra de la quantité d'information recueillie). Le risque associé à une décision est finalement mesuré en tenant compte du pire choix de distribution dans cet ensemble. Nous démontrons notamment que, pour une grande fourchette de problèmes, si l'information se réduit au soutien, à la moyenne et à la matrice de covariance, la décision optimale peut alors être trouvée très rapidement. Ce résultat peut sembler surprenant, puisque résoudre certains de ces problèmes pour une seule distribution peut s'avérer très difficile.

Nous proposons également une méthode qui peut être utilisée pour déterminer les caractéristiques intéressantes d'une distribution à partir de données historiques. Si chaque échantillon est indépendant et identiquement distribué, et l'ampleur maximale du vecteur de valeurs incertaines n'est pas trop grande, « l'ensemble distributionnel » décrit précédemment a de fortes chances d'être assuré de contenir la vraie distribution associée aux échantillons historiques. Nous sommes donc assurés « probabilistiquement » que le risque mesuré représente une bonne approximation du véritable risque et qu'il n'est pas sous-estimé. En fait, ces outils nous permettent de renoncer à notre foi en un modèle probabiliste subjectif et de miser plutôt sur une théorie des probabilités bien établie.

Cette « approche distributionnellement robuste » a été évaluée à partir d'un problème de sélection de portefeuilles impliquant des données historiques téléchargées du site de *Yahoo! Finances*; le cours de 30 titres financiers fut suivi de 2001 à 2007. Les résultats nous démontrent clairement que, bien qu'une approche fondée sur un modèle de distribution (en particulier, le modèle de distribution empirique) puisse générer des retours positifs, l'exposition au risque peut être réduite de manière significative, sans pertes au plan des retours, en reconnaissant simplement que les distributions sont faillibles. ■

> Afin d'expliquer la surexposition au risque de plusieurs fonds de ce type, nombreux sont ceux qui ont attribué la situation à la confiance aveugle des gestionnaires de portefeuilles en leur modèle stochastique –tenant responsable, par exemple, l'utilisation du modèle log-normal en faveur de la tarification des options ou du modèle de copule gaussienne en faveur des CDO (*Collateralized Debt Obligations* : obligations adossées à des actifs).

Operations Research, 58(3), pp. 596-612, 2010.
Titre original
Distributionally Robust Optimization under Moment Uncertainty with Application to Data-Driven Problems

Erick Delage
Service de l'enseignement des méthodes quantitatives de gestion
HEC Montréal et GERAD

Yinyu Ye
Département des sciences de la gestion et du génie
Université de Stanford, Californie