Avis en ligne : ce que mesure une note moyenne

Tyler Kohler . 10 novembre 2026 . note 07

RESUME

La note moyenne affichée sur une fiche est lue comme une mesure de qualité. Cette note examine ce qu’elle agrège réellement. J’y défends l’hypothèse qu’elle est le produit de quatre filtres successifs, qui portent sur les personnes qui achètent, sur celles qui écrivent, sur ce que les avis précédents font écrire et sur ceux qui trichent, puis d’un arrondi, et qu’elle agit sur la demande par seuils. Le mécanisme est appuyé sur les travaux consacrés à l’autosélection des auteurs d’avis, sur une expérience randomisée d’influence sociale, sur les études de la fraude et sur deux estimations par discontinuité de régression. Je propose d’appeler avis pivot l’avis qui fait franchir un seuil d’arrondi à la moyenne affichée. Les limites sont exposées.

Note moyenne affichée
Moyenne arithmétique des notes déposées sur une fiche, arrondie selon la règle de la plateforme avant d’être montrée. Le public lit la valeur arrondie.

Autosélection
Fait que les clients qui déposent un avis diffèrent de ceux qui n’en déposent pas. Hu, Pavlou et Zhang (2017) distinguent le biais d’acquisition, qui porte sur les personnes qui achètent, et le biais de sous-déclaration, qui porte sur celles qui écrivent.

Distribution en J
Répartition des notes observée sur de nombreuses plateformes : une majorité de notes maximales, une minorité de notes minimales, très peu de notes intermédiaires (Hu, Zhang et Pavlou, 2009).

Faux avis
Avis déposé par une personne qui n’a pas eu l’expérience qu’elle décrit, ou qui l’a rédigé sur commande ou contre rémunération, que le but soit de promouvoir une fiche ou d’en dégrader une autre.

Akerlof (1970) a décrit ce qui arrive à un marché où le vendeur connaît la qualité de ce qu’il vend et où l’acheteur l’ignore : les bons produits, impossibles à distinguer des mauvais, cessent d’y être proposés. Les avis en ligne passent pour avoir résolu ce problème à grande échelle. Chaque acheteur dépose son expérience, et la plateforme en tire un nombre.

Ce nombre tient en deux caractères sur une fiche : 4,3. Il est lu comme une mesure de qualité, parfois comme la seule. La question de ce qu’il mesure a été abondamment étudiée, et la réponse publiée s’écarte de l’usage qu’on en fait.

L’hypothèse et ses conditions de validité

Hypothèse : la note moyenne affichée est le produit de quatre filtres successifs, suivis d’un arrondi. Les filtres portent sur les personnes qui achètent, sur celles qui écrivent, sur ce que les avis déjà visibles font écrire aux suivants, et sur ceux qui trichent. L’arrondi transforme ensuite une grandeur continue en paliers, et la demande réagit aux paliers.

L’hypothèse vaut pour des plateformes où le dépôt d’un avis est volontaire, où la moyenne est montrée avant le détail, et où le lecteur n’a pas d’expérience directe du vendeur. Elle perd de sa force quand le volume d’avis est très élevé, quand la plateforme sollicite systématiquement tous les acheteurs, ou quand le lecteur dispose d’une autre source de réputation. Luca (2011) le constate pour les chaînes de restaurants, dont la demande ne réagit pas à la note.

Elle serait affaiblie par la démonstration que les moyennes convergent vers une mesure indépendante de la qualité à mesure que les avis s’accumulent. Le travail qui s’en approche le plus, discuté plus bas, ne trouve pas cette convergence.

Quatre termes définis avant usage

Le premier terme sépare ce que la plateforme calcule de ce que le public lit.

Le deuxième décrit la population dont la moyenne est tirée.

Le troisième décrit la forme que prend cette population une fois comptée.

Le dernier est employé dans un sens large par la presse et dans un sens étroit par les plateformes. Je retiens une définition par l’expérience et par la commande.

Le mécanisme proposé

Premier filtre, l’acquisition. N’achète que celui qui s’attend à être satisfait. La population des acheteurs est donc déjà orientée avant le premier avis. Li et Hitt (2008) en tirent un modèle, qu’ils testent sur des avis de livres : les premiers acheteurs ont des préférences propres, leurs notes diffèrent de celles que donneront les suivants, et les lecteurs de ces premiers avis ne corrigent pas complètement l’écart.

Deuxième filtre, la sous-déclaration. Parmi les acheteurs, ceux dont l’expérience est extrême écrivent davantage. Hu, Zhang et Pavlou (2009) relèvent, sur des données collectées en 2005 sur Amazon, que 78 % des notes de livres, 73 % des notes de DVD et 72 % des notes de vidéos valent quatre étoiles ou plus, avec une seconde concentration à une étoile et presque rien entre les deux. Les auteurs rappellent qu’une moyenne ne résume correctement une distribution que si celle-ci présente un seul sommet ou deux sommets symétriques. Une distribution en J ne remplit aucune des deux conditions.

Schoenmueller, Netzer et Stahl (2020) retrouvent cette polarité sur plus de 280 millions d’avis issus de 25 plateformes. Elle varie avec la sélectivité des auteurs : moins les clients d’une plateforme sont nombreux à écrire, plus la distribution est polarisée, et moins les avis informent. Dellarocas et Wood (2008) avaient traité le silence comme une donnée, sur les évaluations croisées d’eBay : les utilisateurs évaluent plus volontiers une transaction satisfaisante, et le taux de satisfaction affiché, supérieur à 99 %, surestime la qualité réelle des transactions.

Troisième filtre, l’influence sociale. Muchnik, Aral et Taylor (2013) ont attribué au hasard un premier vote, positif ou négatif, à des commentaires publiés sur un site d’agrégation d’actualités. Un premier vote positif accroît de 32 % la probabilité que le vote suivant soit positif, et élève la note finale de 25 % en moyenne. Un premier vote négatif est corrigé par les votants suivants. L’effet est donc asymétrique : la manipulation favorable s’accumule, la manipulation défavorable se résorbe. Le site étudié recueille des votes sur des commentaires, avec une tendance naturelle au vote positif. Une fiche d’établissement n’y est pas assimilable sans précaution.

Godes et Silva (2012) observent un mouvement inverse sur des avis de livres : à date, auteur et ouvrage contrôlés, chaque note est en moyenne inférieure à la précédente. Les deux résultats ne se contredisent pas, puisqu’ils portent sur des objets différents. Ils établissent l’un et l’autre qu’une note dépend de celles qui la précèdent.

Quatrième filtre, la manipulation. Mayzlin, Dover et Chevalier (2014) comparent les avis d’un même hôtel sur deux sites, l’un réservé aux clients ayant séjourné, l’autre ouvert à tous. Les hôtels

indépendants à propriétaire unique, qui ont le plus à gagner à la fraude, ont davantage d'avis positifs sur le site ouvert, et leurs voisins y ont davantage d'avis négatifs. Luca et Zervas (2016) analysent les avis de restaurants que l'algorithme de Yelp écarte comme suspects : environ 16 % du total, plus extrêmes que les autres dans les deux sens. Un restaurant fraude davantage lorsque sa réputation est faible, parce qu'il a peu d'avis ou vient d'en recevoir de mauvais, et il reçoit davantage de faux avis défavorables lorsque la concurrence s'intensifie.

He, Hollenbeck et Proserpio (2022) décrivent le marché lui-même, sur Amazon. Des vendeurs achètent des avis dans de grands groupes privés, y compris pour des produits déjà bien notés. La note et le nombre d'avis augmentent pendant la période d'achat. Lorsque l'achat cesse, la note redescend et la part des avis à une étoile augmente nettement, surtout pour les produits récents.

Reste l'arrondi. Luca (2011) exploite le fait que Yelp arrondit la moyenne au demi-point : deux restaurants dont les moyennes réelles sont presque identiques, l'une juste au-dessous d'un seuil, l'autre juste au-dessus, sont affichés avec une demi-étoile d'écart. En rapprochant les notes des recettes déclarées à l'administration fiscale de l'État de Washington, il estime qu'une étoile affichée supplémentaire se traduit par 5 à 9 % de chiffre d'affaires en plus pour un restaurant indépendant. Le travail est un document de recherche, non publié dans une revue à comité de lecture.

Anderson et Magruder (2012) appliquent la même méthode aux réservations : une demi-étoile affichée de plus conduit un restaurant à afficher complet 19 points de pourcentage plus souvent, soit une hausse de 49 %, et davantage lorsque les autres sources d'information sont rares. Dans les deux cas, la qualité réelle est identique de part et d'autre du seuil. Seul l'affichage diffère.

Ce que montre la littérature, et à quelles conditions

Sur la validité de la moyenne, le travail de référence est celui de de Langhe, Fernbach et Lichtenstein (2016). Sur 1'272 produits de 120 catégories, les notes moyennes des utilisateurs convergent mal avec les scores d'un organisme d'essais indépendant, reposent souvent sur des effectifs insuffisants, ne prédisent pas le prix de revente, et sont plus élevées pour les produits chers et les marques haut de gamme à qualité mesurée égale. Les consommateurs, eux, accordent à la moyenne plus de poids qu'au prix ou au nombre d'avis, et ne modulent pas leur confiance selon l'effectif. Les auteurs parlent d'illusion de validité. Leur corpus est fait de biens de consommation testables en laboratoire. Un service, dont la qualité varie d'une prestation à l'autre, pose un problème différent.

Sur le poids relatif des avis, Chevalier et Mayzlin (2006) comparent les ventes des mêmes livres sur deux sites. Une amélioration des avis sur un site y accroît les ventes relatives, l'effet d'un avis à une étoile dépasse celui d'un avis à cinq étoiles dans la plupart de leurs échantillons, et la longueur des avis indique que les clients lisent le texte. Resnick et coauteurs (2006), dans une expérience de terrain sur eBay, mesurent à 8,1 % du prix de vente l'avantage d'une identité établie sur une identité nouvelle. Un ou deux avis négatifs sur un vendeur récent n'y modifient pas le prix obtenu.

Sur la réponse du professionnel, deux travaux portant sur l'hôtellerie arrivent à des résultats distincts. Proserpio et Zervas (2017) estiment qu'un hôtel qui se met à répondre gagne 0,12 étoile et 12 % de volume d'avis, et qu'il reçoit des avis négatifs moins nombreux et plus longs : un client mécontent hésiterait à laisser un avis bref et indéfendable s'il s'attend à être lu. Chevalier, Dover et Mayzlin (2018) observent que la réponse stimule le dépôt d'avis, et particulièrement d'avis négatifs, le client y voyant le signe que sa plainte sera entendue. Les deux résultats se recoupent sur le volume et divergent sur la tonalité.

Un concept proposé : l'avis pivot

Les deux estimations par discontinuité établissent qu'autour d'un seuil, une variation infime de la moyenne produit un écart de demande mesurable. Elles ne disent rien de l'avis qui provoque cette variation. Je propose de le nommer.

J'appelle avis pivot l'avis dont l'ajout ou le retrait fait franchir à la moyenne affichée un seuil d'arrondi. Soit une fiche de trente avis dont les notes totalisent 127 points : sa moyenne est de 4,23, affichée 4,2 sur une plateforme qui arrondit à la décimale. Un avis supplémentaire à une étoile la porte à 4,13, affichée 4,1. Le retrait d'un avis à une étoile la porte à 4,34, affichée 4,3. Sur une fiche de trois cents avis de même moyenne, aucune de ces deux opérations ne modifie l'affichage.

Le schéma reprend cet exemple : la moyenne réelle varie de façon continue, la note affichée par paliers, et un seul avis suffit à changer de palier quand le volume est faible.

Le pivot dépend donc de deux grandeurs : le volume d'avis, qui fixe le poids de chacun, et la distance de la moyenne au seuil le plus proche. Sur une fiche à faible volume, presque chaque avis extrême est pivot. Sur une fiche à fort volume, aucun ne l'est isolément.

Ce concept est une hypothèse, et sa portée dépend d'un point que les travaux cités ne tranchent pas. Les discontinuités mesurées portent sur un arrondi au demi-point. Beaucoup de plateformes, dont Google au moment où j'écris, affichent une décimale. Le pas y est cinq fois plus fin, et je ne connais aucune estimation de l'effet d'un dixième de point sur la demande. Il est possible qu'il soit négligeable, ou qu'il ne joue qu'à certains paliers, par exemple au passage sous une valeur entière. Le concept deviendrait réfutable par une discontinuité de régression conduite à chaque dixième.

Limites et zones d'incertitude

Les données viennent presque toutes des États-Unis, de quatre ou cinq plateformes et de trois secteurs : la restauration, l'hôtellerie, les biens culturels. Les services professionnels, la santé, le commerce entre entreprises n'y figurent pas. Le recours aux avis diffère probablement quand la décision engage davantage.

Les estimations de fraude reposent sur des approximations. Un avis écarté par l'algorithme de Yelp n'est pas un faux avis établi. Luca et Zervas le traitent comme un indicateur et s'en expliquent. L'écart entre deux sites mesuré par Mayzlin et coauteurs révèle une fraude d'ensemble. Il ne désigne aucun avis en particulier.

Le modèle des quatre filtres présente la moyenne comme une grandeur biaisée dans un sens connu. Ce n'est pas le cas. Les filtres jouent en sens opposés : l'acquisition et l'influence sociale tirent vers le haut, la sous-déclaration gonfle les deux extrêmes, la séquence tire vers le bas. Leur résultante sur une fiche donnée n'est pas calculable à partir de la fiche.

Hu, Pavlou et Zhang (2017) montrent enfin que les consommateurs perçoivent en partie ces biais et tentent de les corriger à l'aide d'autres paramètres de la distribution. La correction reste incomplète. Elle existe, et le lecteur d'avis n'est pas aussi naïf que le suggère une lecture rapide de ces travaux.

Implications

Ce qui suit hérite des limites exposées plus haut.

Pour qui lit une fiche, la moyenne est l'information la moins fiable de la page. Le volume, la forme de la distribution, la date des avis et leur texte renseignent davantage, et la littérature indique que le

texte est lu.

Pour un établissement, le volume commande l'exposition. Une fiche de quelques dizaines d'avis est à la merci d'un seul, authentique ou non. La même fiche à plusieurs centaines d'avis devient à peu près insensible à l'avis isolé. Un recueil régulier auprès de l'ensemble des clients, sans tri préalable des satisfaits, réduit l'autosélection et le nombre d'avis pivots.

Un faux avis défavorable n'a pas le même poids partout. D'après Luca et Zervas, il apparaît surtout là où la concurrence est vive. D'après l'arithmétique du pivot, il pèse surtout sur les fiches à faible volume. Les deux conditions sont souvent réunies chez un indépendant récent, qui est aussi, selon Luca, celui dont le chiffre d'affaires dépend le plus de la note.

L'achat d'avis favorables produit, d'après He et coauteurs, une hausse limitée à la période d'achat, suivie d'une baisse de la note et d'une montée des avis à une étoile. Les auteurs y voient l'indice que ces avis trompent les acheteurs sur la qualité du produit.

Répondre aux avis modifie ce que les clients suivants écrivent. Les deux travaux disponibles s'accordent sur ce point et divergent sur le sens du changement, ce qui interdit d'en faire une règle.

La question de savoir si une vague d'avis est spontanée ou organisée relève d'un autre examen, décrit dans la note sur ce qui distingue une campagne d'une colère. Le coût de production d'un avis plausible, devenu presque nul, est traité dans la note sur l'écart entre produire et vérifier.

Deux fiches affichant 4,3 peuvent décrire deux réalités sans rapport : vingt-cinq notes maximales et cinq notes minimales pour l'une, deux cent dix notes de quatre étoiles et quatre-vingt-dix de cinq pour l'autre. Le nombre affiché est le même.

REFERENCES

AKERLOF, George A (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. The Quarterly Journal of Economics, vol. 84, no 3, p. 488-500.

CHEVALIER, Judith A. et MAYZLIN, Dina (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. Journal of Marketing Research, vol. 43, no 3, p. 345-354.

RESNICK, Paul, ZECKHAUSER, Richard, SWANSON, John et LOCKWOOD, Kate (2006). The value of reputation on eBay: A controlled experiment. Experimental Economics, vol. 9, no 2, p. 79-101.

DELLAROCAS, Chrysanthos et WOOD, Charles A (2008). The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. Management Science, vol. 54, no 3, p. 460-476.

LI, Xinxin et HITT, Lorin M (2008). Self-Selection and Information Role of Online Product Reviews. Information Systems Research, vol. 19, no 4, p. 456-474.

HU, Nan, ZHANG, Jie et PAVLOU, Paul A (2009). Overcoming the J-shaped distribution of product reviews. Communications of the ACM, vol. 52, no 10, p. 144-147.

LUCA, Michael (2011). Reviews, Reputation, and Revenue: The Case of Yelp.com. Harvard Business School, document de travail no 12-016 (non publie en revue).

ANDERSON, Michael et MAGRUDER, Jeremy (2012). Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database. The Economic Journal, vol. 122, no 563, p. 957-989.

GODES, David et SILVA, José C (2012). Sequential and Temporal Dynamics of Online Opinion. Marketing Science, vol. 31, no 3, p. 448-473.

MUCHNIK, Lev, ARAL, Sinan et TAYLOR, Sean J (2013). Social Influence Bias: A Randomized Experiment. Science, vol. 341, no 6146, p. 647-651.

MAYZLIN, Dina, DOVER, Yaniv et CHEVALIER, Judith (2014). Promotional Reviews: An Empirical Investigation of

Online Review Manipulation. American Economic Review, vol. 104, no 8, p. 2421-2455.

DE LANGHE, Bart, FERNBACH, Philip M. et LICHTENSTEIN, Donald R (2016). Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings. Journal of Consumer Research, vol. 42, no 6, p. 817-833.

LUCA, Michael et ZERVAS, Georgios (2016). Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud. Management Science, vol. 62, no 12, p. 3412-3427.

HU, Nan, PAVLOU, Paul A. et ZHANG, Jie (2017). On Self-Selection Biases in Online Product Reviews. MIS Quarterly, vol. 41, no 2, p. 449-471.

PROSERPIO, Davide et ZERVAS, Georgios (2017). Online Reputation Management: Estimating the Impact of Management Responses on Consumer Reviews. Marketing Science, vol. 36, no 5, p. 645-665.

CHEVALIER, Judith A., DOVER, Yaniv et MAYZLIN, Dina (2018). Channels of Impact: User Reviews When Quality Is Dynamic and Managers Respond. Marketing Science, vol. 37, no 5, p. 688-709.

SCHOENMUELLER, Verena, NETZER, Oded et STAHL, Florian (2020). The Polarity of Online Reviews: Prevalence, Drivers and Implications. Journal of Marketing Research, vol. 57, no 5, p. 853-877.

HE, Sherry, HOLLENBECK, Brett et PROSERPIO, Davide (2022). The Market for Fake Reviews. Marketing Science, vol. 41, no 5, p. 896-921.

Kohler, T. (2026). Avis en ligne : ce que mesure une note moyenne. Notes, tylerkohler.ch. https://tylerkohler.ch/notes/avis-en-ligne/