Intervalle de confiance et test d'hypothèse

Les concepts sur un exemple paradigmatique

Marc Bourdeau

There is no true value of anything.

W. E. Deming [1900-1993]

Les quantités exactes n'existent pas, qui en doute? Attention, il s'agit bien de quantité et non de qualité. Elles sont toutes sujettes à erreurs de mesure, imprécisions, aléas. On mesure 10 fois une quantité, on obtient 10 mesures différentes, malgré toutes les précautions pour éviter les erreurs de mesure… à condition d'avoir un instrument assez précis. Ce qui existe au fond, c'est la *répartition des valeurs* sur la droite dont on cherche à préciser certains paramètres.

Elle devrait être unique, on parle de la loi de la mesure. Ainsi, une valeur centrale, une moyenne; une mesure de dispersion, telles un écart type, une étendue. Chaque mesure obtenue est donc une réalisation d'une variable dite aléatoire (VA). Les paramètres de ces VA sont en quelque sorte virtuels, on ne peut les observer directement. C'est ainsi qu'on est amené à développer le concept d'*intervalle de confiance* pour un paramètre (existe-t-il une autre façon?).

Pour comprendre concrètement le concept d'intervalle de confiance, rien de tel que de recourir à un exemple paradigmatique, tous les autres se conjuguent sur le même modèle, *mutatis mutandis.*

Disons qu'on cherche à estimer la valeur moyenne, μ, d'une VA (mesure d'une quantité) X. On peut échantillonner n fois les réalisations de X. On dit de façon indépendantes d'un échantillon à l'autre : on ne développe pas ici le concept d'indépendance, qu'on laisse à l'intuition, de même que le concept précis de VA.

On sait que *Xbar*, la moyenne des échantillons a la même moyenne que celle X, la mesure elle-même, mais un écart type, $\sigma/\sqrt{n}$, où σ est l'écart type de X. Donc plus n augmente plus l'écart type de *Xbar* diminue. Ainsi il est divisé par deux lorsqu'on quadruple la taille n de l'échantillon. Notons aussi que *Xbar* est une VA, si on change d'échantillons on change la valeur de *Xbar*. Plus on échantillonne X, plus n est élevé plus les chances augmentent que X*bar,* soit près de μ, l'écart type de *Xbar,* diminuant au fur et à mesure de n augmente.

L'important ici est qu'on fait jouer le résultat le plus important de la probabilité, sur lequel repose presque toute la statistique, le bien nommé Théorème central de la limite, le TCL : la loi de répartition de *Xbar,* on dit simplement la loi de *Xbar* est gaussienne ('normale') dès que n est assez grand, ou que la loi de X est elle-même gaussienne et ce quel que soit n.[1]

Si on ne connaît pas σ, ce qui est le cas habituel, on estime σ par s, l'écart type échantillonnal de X, on obtient alors une loi T (loi de Student) plutôt que la gaussienne pour *Xbar*, elle-même tabulée comme la gaussienne. Ce qui, en pratique, ne change presque rien, dès que n est assez grand (disons $n \geq 60$).

Admettons donc que *Xbar* est gaussienne. On sait calculer, pour toute couverture p désirée, les valeurs autour de la moyenne μ assurés de couvrir les valeurs de *Xbar*, puisqu'on connaît sa loi, centrée en μ et d'écart type $s/\sqrt{n}$. Les valeurs de cet intervalle dit de confiance, calculées sans aucun paramètre autre que ceux fournis par l'échantillon, notées *IC(p)*, pour tout p, font jouer les quantiles de la loi gaussienne normée (de moyenne 0 et d'écart type unité), les valeurs de Xbar, de $s/\sqrt{n}$. *Xbar*, qui est dans ce cas la valeur centrale de l'intervalle *IC(p)*. La formule pour déterminer cet intervalle, quoique simple, n'est pas intéressante pour comprendre le concept. On obtient alors que μ est située dans son *IC(p)* avec une *confiance p.*

Attention ici, le concept n'est pas aussi simple qu'il n'y paraît. La raison pour laquelle on parle de *confiance* et non de *probabilité*, est que nous avons affaire à un intervalle probabiliste, au sens où il est aléatoire lui-même : changeons les échantillons de X, ne fût-ce que d'un seul, on change les limites de *IC(P)*. On sait alors qu'en *moyenne*[2] les *IC(p)* contiennent μ 100p% des fois et ne contiennent pas μ dans 100(1-p)%, sur 100 réalisations de *IP(p)* (il s'agit de la moyenne d'une loi binomiale *Bi(n,p)*). Par exemple, si on examine 100 échantillons de taille n de la loi de X, et qu'on calcule les 100 *IP(p=0,95)*, on en trouvera en *moyenne* 5 qui ne contiendront pas le μ recherché. C'est un abus de langage de dire qu'un *IC(p)* particulier a une *probabilité p* de contenir le paramètre recherché, ici μ. Les étudiants, ainsi que bien des statisticiens de fortune, même Wikipedia! comprennent cela avec difficulté…

C'est ainsi qu'on annonce, par exemple, que les intentions de vote pour un candidat, obtenu d'un sondage de taille *n=1000,* est un « certain % (la proportion échantillonnale en % calculée sur l'échantillon) plus ou moins 3%, 95% des fois » (empilade de pourcentages un peu sibyllisn !), c'est l'intervalle de confiance de *niveau* 95%. Encore une fois, au sens où sur 100 échantillons de cette mesure (prises au même moment, indépendants et de même taille), on aura en moyenne que 95% contiendront la vraie valeur des intentions de vote (au moment de la réalisation du sondage, c'est entendu…). Rien n'est précisé pour un seul sondage… Et notons que pour réduire la marge d'erreur de 3 à 1,5%, au même 95% de confiance par un facteur 2 donc, il faudra multiplier la taille échantillonnale par 4…

---

[1] Ce résultat tient pour une vaste classe de lois aussi peu symétriques fussent-elles, pourvu qu'elles soient de moyenne et variance finies. Plus elles sont dissymétriques, plus la taille de l'échantillon doit être élevée : $n \geq 10$ suffit pour les lois assez symétriques; avec $n \geq 30$, on peut être assuré du résultat dans presque tous les cas pratiques.

[2] « *On the long run* », comme on dit en anglais, en moyenne sur un grand nombre d'échantillons de taille n donc.

Notons enfin qu'on a une probabilité nulle que ce pourcentage réel soit exactement celui obtenu de l'échantillon, *IC(p)* étant réduit à la valeur de *Xbar,* et *p*=0%, mais il est de p=100% que celui de l'échantillon soit quelque part sur toute la droite réelle!

Dans les deux cas aucune information intéressante n'est obtenue d'un échantillon. Il faut se donner un *p* disons raisonnable… On ne développera pas non plus ici ce point délicat. 95% est le *p conventionnel.* Le lecteur intéressé pourra verbaliser le concept pour *p=50%, p= 99%...*

Par ailleurs, si on pense à une certaine valeur pour μ, disons μ_0, et que cette valeur hypothétique n'est pas située dans *IC(p)*, qui est l'intervalle de confiance pour μ, la valeur μ_0 est située dans le complémentaire de la couverture de probabilité *p*, soit donc *IC(p)*, c'est ce qu'on appelle la *région de rareté 1-p* pour le paramètre μ. Ainsi par exemple pour *p = 0,95, 1-p = 0,05*, et il y a peu de chance, dans ce cas, que μ_0 soit une valeur probable pour μ. On ne parierait pas aisément sur la base de l'information fournie par l'échantillon, à moins d'être un mauvais parieur…, que la valeur hypothétique μ_0 est la vraie valeur de μ. En termes intuitifs : la valeur présumée de μ_0 n'est pas dans la zone de p-couverture de μ défini par son intervalle de confiance *IC(p)* autour de *Xbar*, elle est dans une zone de rareté *(1-p)* pour μ. Pour p=0,95 par exemple, 1-p=0,05 : qui parierait sur un événement qui n'a qu'une chance sur 20 de se réaliser? Tout parieur comprend qu'on peut le faire à une forte cote de 1 contre 20, i.e. gagner c'est gagner 20 fois la mise. Le problème est de trouver un contre-parieur peu conscient de son risque… C'est là en effet l'art de gagner ses paris, convaincre/tromper…

Il y a toutes sortes de détails sur cette théorie, notamment les intervalles de confiance unilatéraux, qu'on ne développera pas ici. Le lecteur intéressé pourra trouver cependant, un développement, lui aussi presque sans équations, du paradigme classique, ou de Neyman-Pearson, des tests d'hypothèse.

Le concept des *IC(p)* est-il devenu accessible, clair, incompréhensible ? Ces explications sont peut-être trop succinctes… À vous de me le signifier !

— 13 avril 2017