Théorie de la réponse à l'item

(IRT)

Objectif

L'IRT est un modèle statistique qui permet d'évaluer des corrélations entre les compétences d'un individu, la difficulté d'un item dans un test et sa probabilité d'y répondre correctement. Ce modèle prend en compte plusieurs paramètres liés aux individus ou au test. Il permet entre autre d'établir des algorithmes de proposition d'exercices adaptés à des individus en fonction de l'estimation de leurs compétences au travers de leurs réponses aux items.

Construction

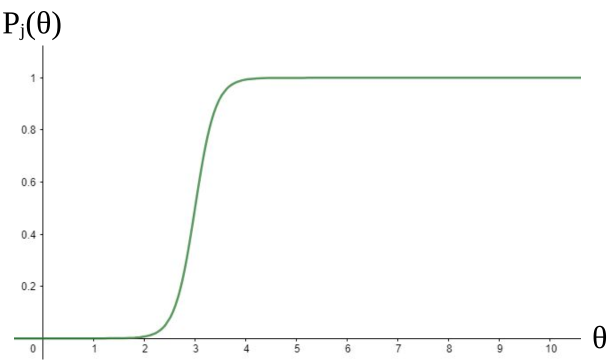
Postulats :

On suppose que la probabilité d'une réponse correcte est attribuable à la position du sujet sur un nombre spécifique k de traits latents de compétence requise pour répondre au type d'items en question.

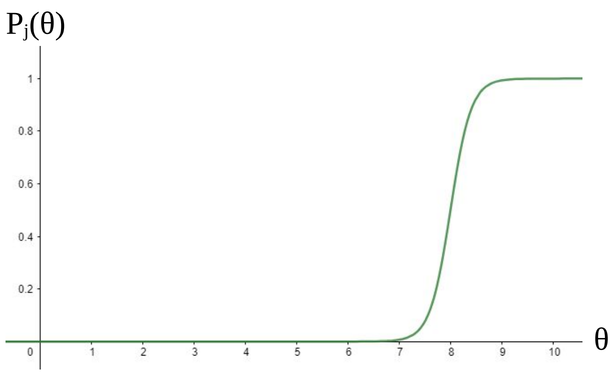
On suppose qu'il y a indépendance locale, c'est-à-dire que la probabilité d'une réponse correcte d'un sujet à un item donné ne dépend pas de ses réponses aux autres items du test.

La probabilité de réponse correcte à un item j mobilisant la compétence θ est modélisé par la fonction de réponse à l'item (IRF) suivante :

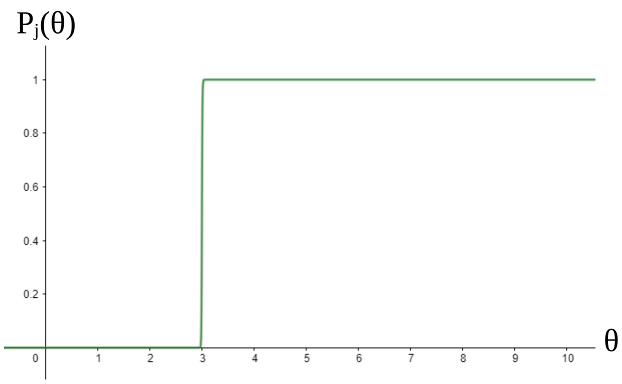
$P_j(\theta)=\frac{1}{1+e^{-a_j(\theta-b_j)}}$ a_j désigne la discrimation de l'item et b_j désigne la difficulté de l'item



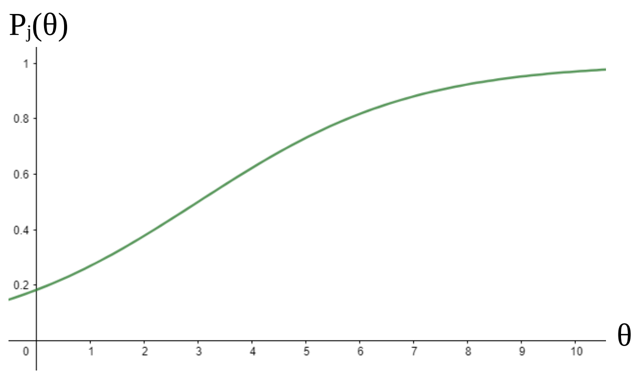
Fonction de réponse à l'item avec aj=5 et bj=3.



Fonction de réponse à l'item avec aj=5 et bj=7.



Fonction de réponse à l'item avec aj=200 et bj=3.
La discrimation de l'item est trés forte, l'individu doit posséder indispensablement la compétence 3 pour répondre correctement.



Fonction de réponse à l'item avec aj=0,5 et bj=3.
La discrimination de l'item est faible, l'individu a des chances de répondre correctement sans requérir la compétence 3. De plus, la compétence 3 n'assure pas entièrement la réponse correcte de l'individu.

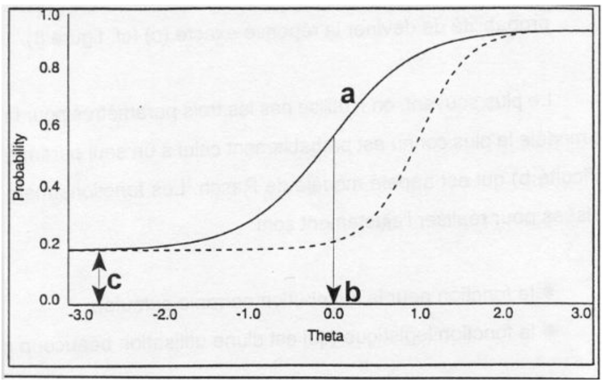


Figure extraite des diapositives de Jang Schiltz, voir annexe.

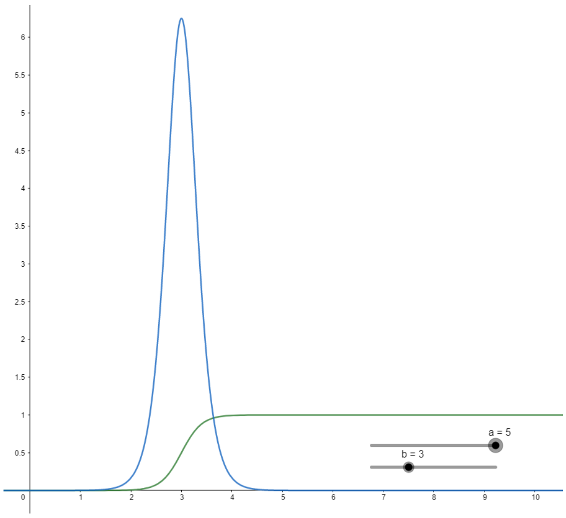
Graphiquement le paramètre a ou a_j caractérise la pente de l'IRF et le paramètre b ou b_j détermine l'abscisse dont l'image correspond à la probabilité de réponse correcte de 50%.

Il est aussi possible d'ajouter un troisième paramètre c dit de "guessing" modélisant la probabilité que l'individu donne la réponse exacte par hasard.

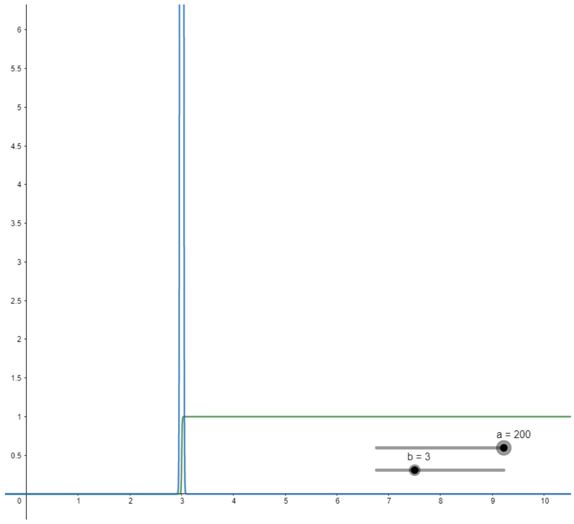
L'IRT fournit aussi la modélisation de la densité de compétence évalué par item grâce à la fonction d'information de l'item (IIF) suivante :

$$I(\theta)=a_j^2 P_j(\theta)*(1-P_j(\theta))$$

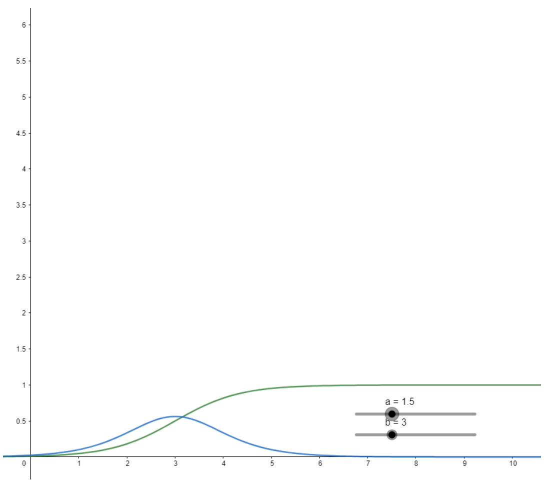
L'IIF permet d'associer à une IRF la précision d'évaluation des compétences évaluées.



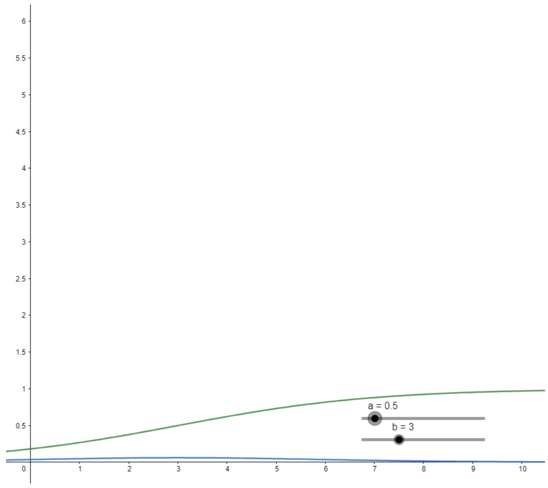
IIF de l'item avec un fort paramètre de discrimination, cette item cible davantage la compétence 3 que les autres.



IIF de l'item avec un trés fort paramètre de discrimination, cette item cible exclusivement la compétence 3.

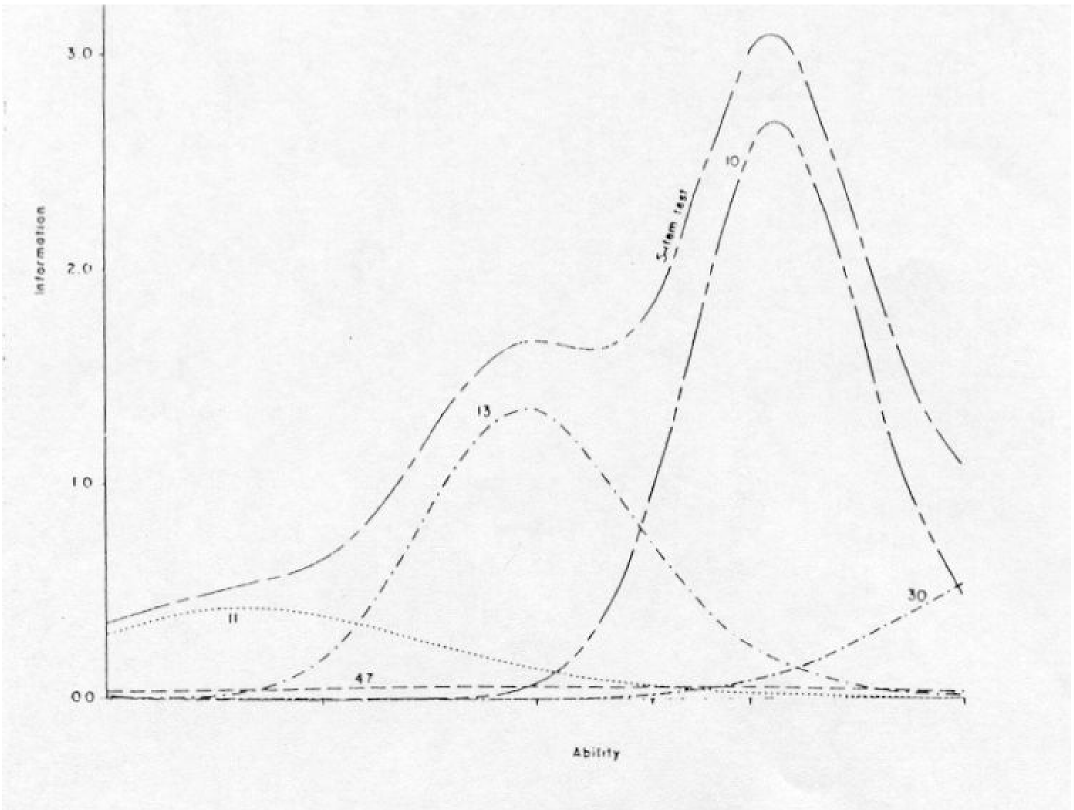


IIF de l'item avec un faible paramètre de discrimination, cette item cible sensiblement la compétence 3 plûtot que les autres.



IIF de l'item avec un trés faible paramètre de discrimination, cette item cible indifféremment toutes les compétences.

Le postulat d'indépendance locale permet la sommation de l'information des différents items pour un niveau de compétence donné. On obtient ainsi la fonction d'information du test (TIF) étant égale à la somme de toutes les (IIT) du test.



Fonction d'information du test (TIF), figure extraite des diapositives de Jang Schiltz, voir annexe

La TIF caractérise la distribution des compétences évaluées par l'ensemble des items, donc du test entier. Elle permet de savoir si le test privilégie certaines compétences et peut servir d'outil pour équilibrer le test en remplaçant ou en ajoutant des items visant les compétences dont le test n'atteste pas beaucoup d'information.

Il est ensuite possible d'effectuer une estimation du niveau de compétence par deux techniques différentes. La première méthode courante est d'effectuer l'estimation du maximum de vraisemblance la deuxième utilise l'estimation modale bayésienne.

L'estimation du maximum de vraisemblance des réponses du sujet i est la valeur θ qui maximise la fonction L_i déterminée par $\log L_i(\theta)=\sum_j \{x_{ij} \log\ P_j(\theta)+(1-x_{ij})\log[1-P_j(\theta)]\}$ où x_{ij} désigne le succés de réponse correcte (0 ou 1) du sujet I à l'item j.

Il faut alors résoudre l'équation de vraisemblance suivante

$$\frac{\partial logL_i(\theta)}{\partial\theta}=\sum_j\left(\frac{x_{ij}-P_j(\theta)}{P_j(\theta)[1-P_j(\theta)]}\right)\left(\frac{\partial P_j(\theta)}{\partial\theta}\right)=0$$

La méthode de l'estimation modale bayésienne utilise une démarche quasi identique.
Exemple d'algorithme utilisant la modélisation IRT :

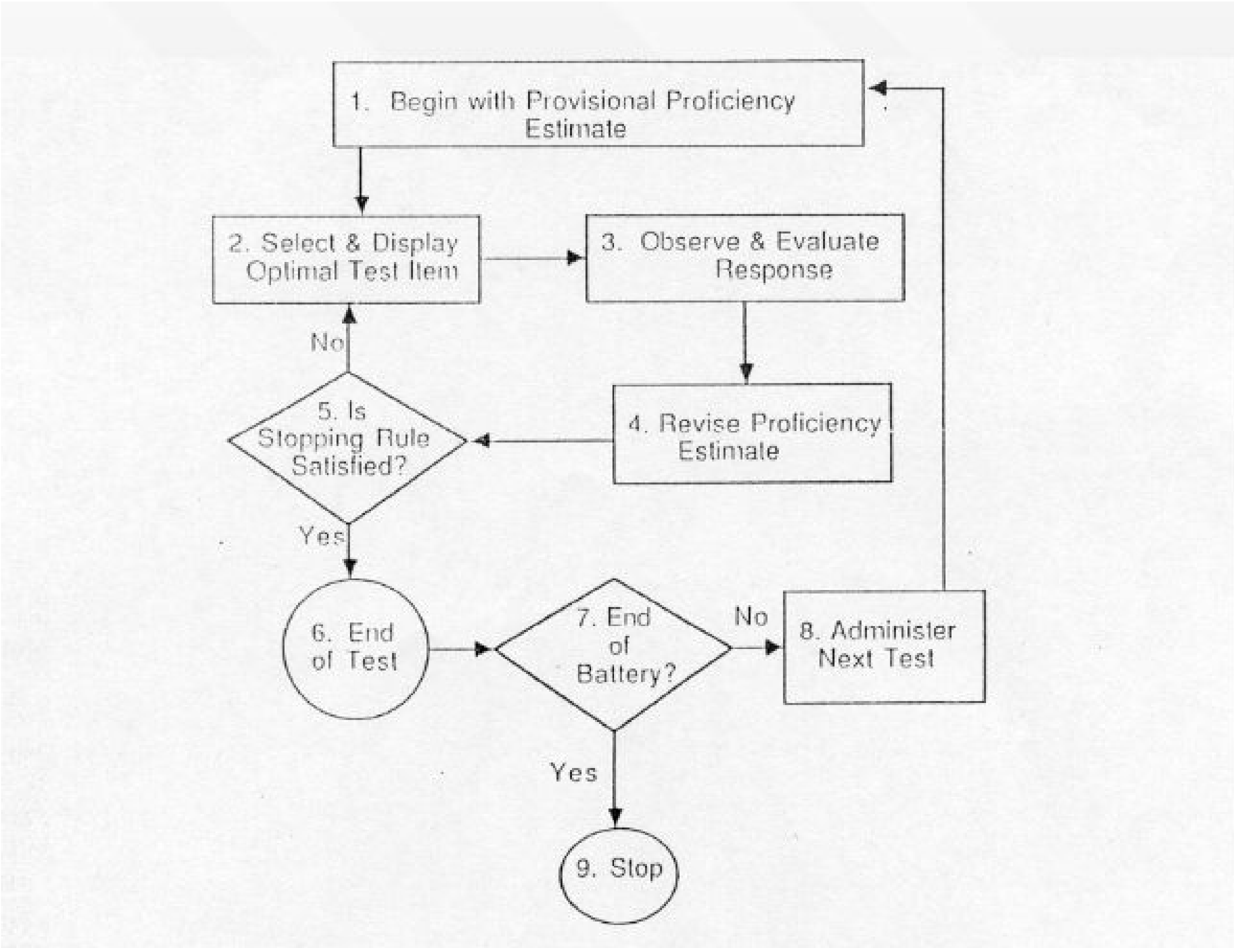


Figure extraite des diapositives de Jang Schiltz, voir annexe

Un test sur mesure peut se terminer quand une précision déterminée est atteinte, mais aussi, après un nombre fixé d'items, ou lorsqu'un temps déterminé s'est écoulé.
En pratique, lorsqu'on veut atteindre un degré de précision déterminé, on impose souvent simultanément un nombre maximal d'items administrés car, dans certains cas, la banque d'items peut être épuisée avant que la précision voulue ne soit atteinte.

Annexe :
"Développement et évaluation d'un test mathématiques adapté, informatisé par ordinateur" de Jang Schiltz :
https://orbilu.uni.lu/bitstream/10993/8720/1/2006_3%20Lille.pdf