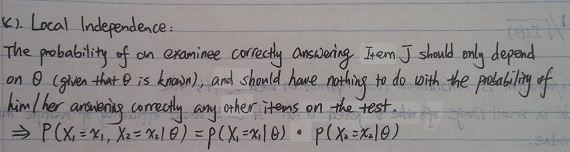
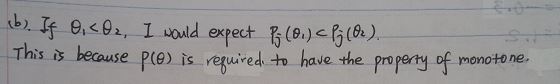
1. The conditional expected score:



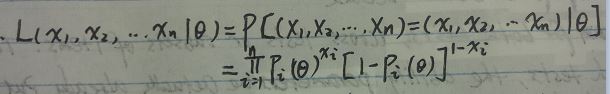
1. The LI assumption is important, because only when this assumption holds, can we attribute completely examinees’ performance to the latent trait(s), instead of to other factors such as latent traits that we don’t care about or characteristics of items.

Also, under the assumption of LI, we are able to easily write down the likelihood function foras the product of , and therefore come up with  estimates.

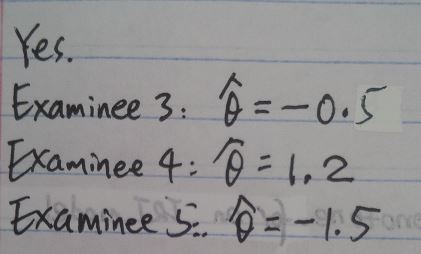
1. Conditional probability



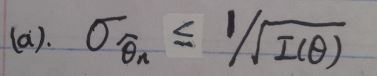
1. Likelihood function



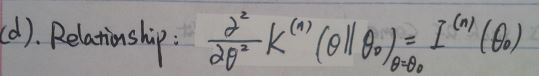
1. Difference: in (a),  is a fixed point, while in (b),  is unknown, and can be estimated by maximizing the likelihood (or log-likelihood) function.
2. The likelihood function is useful in  estimation because we can obtain the likelihood of an examinee having a certain response pattern given various s, and the  that maximizes the likelihood function is the MLE of the true, denoted. By using the likelihood function, we are able to figure out  by simply looking at the graphs of the log-likelihood for each examinee.



* 1. Relationship



* 1. Fisher test information is considered local information because it’s the information within a small range around the targeted. It cares about efficiency of multiple MLEs of, and at, I is a value.
  2. KL information is considered global information because it’s the information outside the small range around the targeted, and it cares about consistency of the estimator. At , KL is a curve instead of a value.
  3. Relationship



* 1. K is more informative than I because by taking derivatives of K, I can be fully recovered, which means knowing K is the same as knowing both K and I. However, knowing I won’t help to obtain K.

1. Yes we can say that Mary outperforms John, because
   1. in IRT, an examinee’s ability parameter is invariant across tests or subsets of items, which means ability estimates, are also invariance across tests or subsets of items. Therefore, in this question, where item parameters are known for both tests, the ability estimates have already been put on the same scale, so they can be compared directly.
   2. <, and s can be compared directly, we can say that Mary outperforms John.
2. Yes I believe that John’s design is reasonable, because
   1. when item parameters are unknown, it is necessary that we remove the indeterminacy in the scale brought by the invariance of the item response function in terms of the linear transformation.
   2. in order to remove the indeterminacy and according to the invariance property, we can either standardize the difficulty parameters or ability parameters, and by fixing the difficulty parameters, we are able to place item parameters estimates and ability estimates on the same scale. However, by standardizing ability values could actually results in item parameters that are on different scales, and therefore it is more appropriate to standardize the b parameters respectively to fix the scale.
3. Item j is biased against the minority group, because after fixing the scale, items parameters are placed on the same scale, and therefore can be compared directly. The fact that >indicates that to have 50% chance of correctly answering item j, minority group members need a larger  than majority group members. Equally, when having the same , minority group members have a lower probability of correctly answering item j compared with majority group members. Therefore, we can say that Item j has DIF, and is biased against the minority group.