1 Neyman-Pearson Classification, NP

For Page-Hinkley Test and other statistics we always need to preselect a threshold λ to balance the trade-off between false positive error and false negative error. While in the Neyman-Pearson classification setting, we could optimize this threshold by introducing the risk to the objective function. Let α be the risk of false positive and β be the risk of false negative, the formulation of NP classification can be summarized as follows:

$$min(\hat{\alpha} - \alpha) + \beta$$

where $\hat{\alpha}$ is a function of λ, i.e. $\hat{\alpha} = f(\lambda)$.

The idea here is by optimizing $min_{\lambda}(f(\lambda) - \alpha) + \beta$, we can select a optimal threshold λ for the trade-off problem. But is this optimization problem convex?

2 Noisy level of users metric, from Michele

For each user we exert a Page-Hinkley Test on its corresponding metric sequence and get a sequence of PH Test values, respectively. Among each PH Test value sequence, there is a set of timestamps whose corresponding PH Test values reveal the potential change points of the metric sequence. We denote $u_{i,i=1,\ldots,n}$ as the set of n users, and $S(u_i)$ be the set of timestamps where each timestamp corresponds to a change in the metric sequence. What we want to verify is:

$$\text{IF } u_i \text{ is less noisy than } u_j\text{, THEN } S(u_i) \subset S(u_j)$$