

EP - Parte 2 - Introdução a Estatística - ACH2053 - 04

Leonardo Kazuhiko Kawazoe 8641959 - Grupo 7

Universidade de São Paulo - Escola de Artes Ciências e Humanidades - São Paulo - SP -
Brazil

3.1 - III1.txt

Para gerar o theta dos 2000 alunos onde as respostas foram informados:

Para cada aluno simulei um range de thetas (Após alguns experimentos decidi que o range ficaria -20.0 e 20.0, onde entre um valor e outro do range existe uma diferença de 0.1).

Range Definido = (-20.0, -19.9, -19.8, ... , 19.9, 20.0)

Para calcular o MLE para cada valor de theta do range verifiquei para cada questão a probabilidade de aquele determinado valor de theta ter obtido aquela resposta (informada no arquivo respostas.txt) para aquela pergunta utilizando TRI. Para resposta dada como correta a probabilidade obtida na função informada no enunciado do EP (Probabilidade de um determinado aluno acertar determinada questão) e para respostas erradas $1 - \text{a probabilidade obtida}$.

Após esse passo fiz o produtório de todas as probabilidades obtidas no passo anterior, descobrindo assim a probabilidade de aquele Theta ter respondido às 100 perguntas definidas com as respostas informadas.

Tendo em mão a distribuição de probabilidades de um certo valor theta ter respondido aquela prova para todos os thetas do range escolhido utilizando uma ordenação dessas probabilidades peguei o theta mais provável de ter respondido aquela prova com aquelas respostas (Maior valor de probabilidade dentro da distribuição) para atribuir como theta daquele aluno. Como desenvolvi em python e criei um objeto exclusivamente para cada ponto dessa distribuição foi relativamente fácil obter o essa ordenação e coletar o máximo absoluto sem precisar fazer uso de um outro método para achar como método da bisseção por exemplo.

Executei o processo algumas vezes e comparei a resposta para verificar se o theta obtido para cada aluno era sempre o mesmo e não existiam falhas no desenvolvimento.

3.2 - II2.txt

Para ver a probabilidade ser do Aluno5 ser melhor que os demais alunos nas provas de tamanho N selecionadas anteriormente e no conjunto de 100 questões apresentadas segui os seguintes passos:

Para cada conjunto de questões simulei a execução da prova 100000 pelos dois alunos comparados utilizando a mesma metodologia usada na parte 1 do EP (Sorteando um valor entre 0 e 1 e verificando se era menor ou igual a probabilidade de um determinado aluno acertar uma determinada questão, a diren).

Para mensurar a habilidade do aluno utilizei TRI da mesma maneira que no exercício anterior, a diferença que dessa vez com o theta já estava definido e não precisamos descobrir o máximo da função mas somente um ponto (Theta informado). Utilizando as respostas (acertos ou erros) simulados, o conjunto de perguntas informado e o theta do aluno cheguei na probabilidade do aluno ter obtido aquelas resposta para aquele conjunto de perguntas.

A probabilidade foi gerada utilizando o confronto entre as vezes que o Aluno5 foi melhor (Verificando se a probabilidade gerada no passo anterior era maior ou igual a do aluno comparado) que o outro aluno avaliado nas 100000 simulações.

0.13263 0.2125 0.34406 0.52029
0.11554 0.24953 0.46775 0.62884
0.05897 0.1915 0.46413 0.67391
0.04626 0.18067 0.47269 0.69314

Podemos verificar que ao contrário do exercício similar na parte 1 do EP aqui as probabilidades aumentam de acordo com o aumento do valor de theta do aluno comparado. O motivo disso é que como usamos TRI na avaliação do aluno e não mais seu número de acertos um aluno com theta alto por exemplo, quanto maior seu theta errando uma questão que é considerada "fácil" tem um redução na sua "nota" de maneira muito drástica causando o comportamento observado.

3.3 - II3.txt

Nesse exercício gerei o intervalo de confiança com grau de confiança $\alpha = 0,1$ aproximando por uma normal para os contextos pedidos no enunciado.

A amostra usada foi obtida utilizando o mesmo processo de avaliação do exercício anterior, através TRI com as respostas (acertos ou erros) simuladas, o conjunto de

perguntas informado e o theta do aluno cheguei na probabilidade do aluno ter obtido aquelas resposta para aquele conjunto de perguntas.

Para o cálculo do intervalo de confiança uma amostra de 100000 resultados foi usada repetindo o experimento citado acima 100000 vezes com cada conjunto de perguntas definido.

O processo repetido para cada aluno versos cada prova de tamanho N definidas anteriormente obtendo o resultado colocado do arquivo de resposta.

3.3 - II4.txt

Nesse exercício gerei o intervalo de confiança com grau de confiança $\alpha = 0,1$ aproximando por uma normal para os contextos pedidos no enunciado.

A amostra de respostas foi obtida da mesma forma que o primeiro exercício da parte 1 do EP. Simulando a execução da prova pelo aluno (Sorteando um valor entre 0 e 1 e verificando se era menor ou igual a probabilidade de um determinado aluno acertar uma determinada questão) para o conjunto de questões informados verificando o número total de acertos.

Para o cálculo do intervalo de confiança uma amostra de 100000 resultados foi usada repetindo o experimento citado acima 100000 vezes com cada conjunto de perguntas definido.

O processo repetido para cada aluno versos cada prova de tamanho N definidas anteriormente obtendo o resultado colocado do arquivo de resposta.