

### 3. Inferência e estimativa de parâmetros (MAP214) (entrega 15 de maio 2003)

24th April 2003

1) Considere uma moeda especial que sera jogada de forma a bater em vários objetos antes de cair. Especial quer dizer que nao temos certeza se ela é honesta, pois foi comprada em um cassino. Introduzimos um parâmetro  $h$  que mede a honestidade de forma que se  $h = 1/2$  a moeda será dita honesta. Se  $h = 0$  só obteremos caras e se  $h = 1$  só obteremos coroas, isto é  $h$  é a probabilidade de que saia uma coroa . Associamos à face uma variável  $s$  que toma valores  $\pm 1$ , i.e  $s_k = 1$  (resp.  $-1$ ) se na  $k$ -ésima jogada o resultado for uma cara (resp. coroa). No arquivo de dados [www.fge.if.usp.br/~nestor/moeda.dat](http://www.fge.if.usp.br/~nestor/moeda.dat) se encontram os resultados de uma sequência de jogadas. Seu objetivo é estimar  $h$ . A distribuição posterior de  $h$  é

$$P(h|D_n, I) = \frac{1}{N_{norm}} P(h, I) P(D_n|h, I)$$

onde  $I$  inclui toda a informação sobre a moeda e a forma como será jogada.  $D_n$  inclui os primeiros  $n$  valores de  $s$  no arquivo de dados. Para a verossimilhança é razoável adotar a forma (não normalizada)

$$P(D_n|h, I) \propto h^l (1-h)^{n-l}.$$

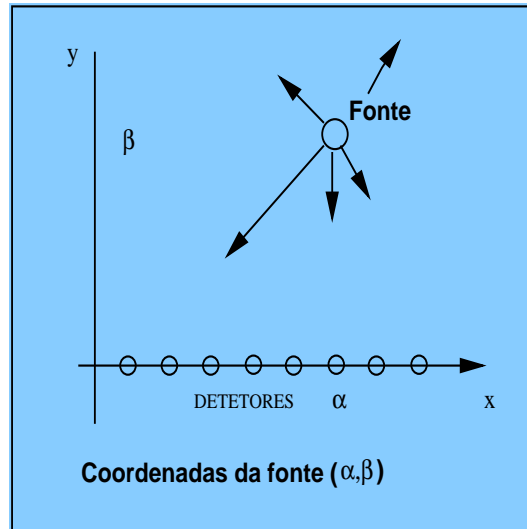
Considere três diferentes possibilidades para a distribuição a priori  $P(h, I)$ : (I) uniforme, que atesta total ignorância antes de ver os dados, (II) proporcional a  $\exp(-(h - .5)^2)/2$  que denota confiança na honestidade da moeda e (III) proporcional a:  $\varepsilon + (x - 0.5)^4$  que demonstra desconfiança,  $\varepsilon = 0.01$ .

(a) Para  $n = 0, 1, 2, 5, 10, 100, 1000, 10000$  jogadas faça o gráfico do logaritmo da posterior como função de  $h$  (apresente estas curvas no mesmo gráfico de forma que o maximo seja igual a 1, i.e. divida todos os pontos pelo valor do máximo). (2ptos)

(b) Do gráfico obtenha o valor mais provável  $h_n$  e a largura a meia altura  $\Delta h_n$  para cada  $n$ . Faça um gráfico log-log. Qual é a lei de evolução de  $\Delta h_n$  com  $n$ ? (2ptos)

2) Considere uma fonte que emite pulsos de luz em direções aleatórias no plano horizontal e um arranjo de sensores dispostos sobre uma reta, como mostra a figura ao lado. A coordenada  $y = \beta = 2$ . é conhecida, enquanto que a abscissa

a posição ao longo do eixo  $x$ , que chamaremos  $\alpha$  não é conhecida e será o objeto de nosso estudo. É dado o conjunto de medidas das posições  $\{x_i\}$  dos sensores que detetaram um pulso no arquivo [www.fge.if.usp.br/~nestor/cauchy.dat](http://www.fge.if.usp.br/~nestor/cauchy.dat).



a única mudança que faremos em relação ao item anterior é que a verossimilhança será

$$\log P(D|\alpha\beta) = \sum_{i=1,\dots,n} \log\left(\frac{1}{\pi} \frac{\beta}{\beta^2 + (x_i - \alpha)^2}\right)$$

(2a) Mostre que esta é uma boa escolha, dado que os ângulos  $\theta_i$  de emissão são independentes, igualmente distribuídos de maneira uniforme entre os valores mínimo e máximo detectáveis  $(-\frac{\pi}{2}, \frac{\pi}{2})$ . (parta de  $P(x)dx = P(\theta)d\theta$ ) (1pto)

(2b) Faça os mesmos gráficos que na primeira parte do problema (log da distribuição posterior). Coloque também no gráfico os valores das médias empíricas de  $x$  dadas por  $\bar{x}_n = \sum_{i=1,\dots,n} x_i/n$ . Note que estas médias não são bons estimadores da posição da fonte. (2Ptos)

(2d) Faça um gráfico da medida empírica da variância  $\sigma_n = \overline{x_n^2} - \bar{x}_n^2$  (1pto)

(2e) Estime as larguras da posterior a meia altura. Note que meia altura numa escala logarítmica quer dizer que você deve procurar os valores da abscissa onde a função tem um valor igual ao do máximo menos  $\log 2$ . (1pto)

(2c) Faça um gráfico da estimativa para  $\alpha$  contra  $n$  baseado no máximo da distribuição posterior com as barras de erro determinadas no item anterior (1pto)