Inferências bayesianas com probabilidade

- Qual é a relação entre inferência bayesiana e as distribuições probabilísticas recém descritas?
- Essa conexão é feita ao se estimar parâmetros da distribuição probabilística usada, ou seja, assume-se um modelo para os dados e tenta-se descobrir as características desse modelo: *modelamento de dados*
- Em outras palavras, tendo uma distribuição probabilística $f(\text{dados}|\vec{\alpha})$ em mente, trata-se de se encontrar o vetor de parâmetros $\vec{\alpha}$.
- O roteiro bayesiano é claro, computar a distribuição posterior de $\vec{\alpha}$ pelo teorema de Bayes, como já fizemos várias vezes.

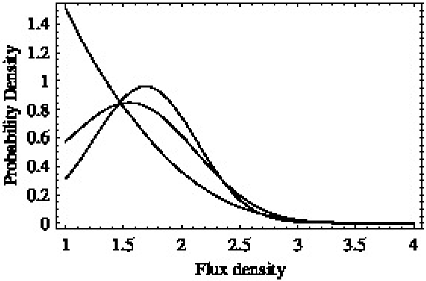
Inferências bayesianas com probabilidade

- Esse método é relacionada com a técnica clássica de *máxima verossimilhança*.
- Se o fator *a prior* é não-informativo ou 'difuso', então o posterior é proporcional ao termo da verossimilhança: $f(\text{dados}|\vec{\alpha})$.
- A máxima verossimilhança toma como resultado os valores de $\vec{\alpha}$ que maximizam a verossimilhança (i.e. a moda do posterior)
- Isso implica em caracterizar o posterior por um único número, o que é útil em várias ocasiões devido a teoremas de máxima verossimilhança (veremos mais sobre isso mais adiante)

Inferências bayesianas com probabilidade

- Seguindo nessa discussão, podemos agora considerar problemas bayesianos.
-

Exemplo W&J P. 47



-
- Nesse exemplo o *a priori* é bem determinado, mas esse não é sempre o caso. Que *a priori* usaríamos caso essa fosse a primeira medida?

Geradores de Distribuição de Monte Carlo: variável aleatória discreta

- Vamos supor que nós queremos calcular os valores da variável aleatória ξ com a seguinte distribuição:

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

- Consideremos agora o intervalo $0 < y < 1$ e dividi-lo em n intervalos com tamanhos iguais a $p_1, p_2, ..., p_n$. As coordenadas das divisões serão então: $y = p_1$, $y = p_1 + p_2$,...,$y = p_1 + p_2 + ..., p_{n-1}$. Numere esses intervalos pelo números 0,1,...,n. (Preparação finalizada; Fig.)

- Para a geração propriamente dita toma-se um valor γ e define-se que $y = \gamma$. Se esse ponto cair no iézimo intervalo, então assume-se que $\xi = x_i$

Geração de uma variável aleatória discreta: algoritmo

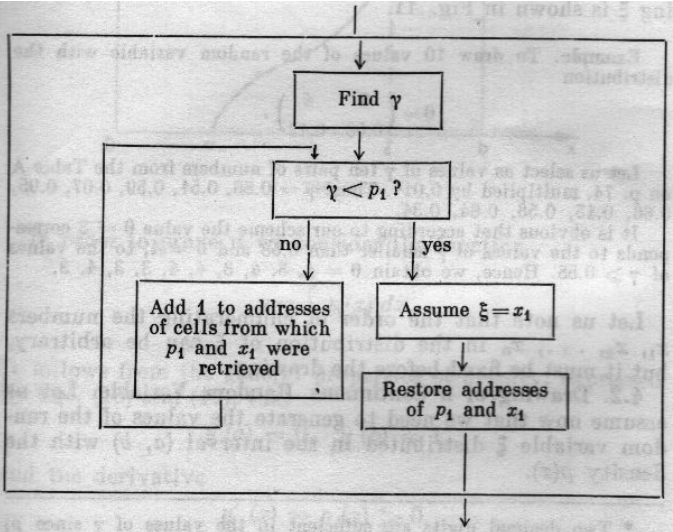


FIG. 11

I. M. Sobol: *The Monte Carlo Method*

Geradores de Distribuição de Monte Carlo: variável aleatória contínua

- Vamos supor agora que queremos gerar os valores de uma variável aleatória ξ distribuída no intervalo (a, b) com densidade $p(x)$.
- Pode ser demonstrado que os valores de ξ podem ser encontrados a partir da equação:

$$\int_a^{\xi} p(x)dx = \gamma \tag{5}$$

- Para provar isso consideremos a função:

$$y = \int_a^x p(x)dx$$

- Das propriedades gerais da PDF decorre que:

$$y(a) = 0, \;\; y(b) = 1,$$

- e que a derivada:

$$y'(x) = p(x) > 0.$$

Geração de uma variável aleatória contínua

- Selecionemos agora um intervalo arbitrário (a', b'), dentro de (a, b), onde
$$y(a') < y < y(b')$$
e
$$a' < x < b'$$
- Consequentemente, se ξ pertence ao intervalo $a' < x < b'$, então γ pertence ao intervalo $y(a') < y < y(b')$, e vice-versa.
- Assim:
$$P\{a' < \xi < b'\} = P\{y(a') < \gamma < y(b')\}$$
- Como γ é uniformemente distribuído em $(0, 1)$,
$$P\{y(a') < \gamma < y(b')\} = y(b') - y(a') = \int_{a'}^{b'} p(x)dx$$
- Assim sendo:
$$P\{a' < \xi < b'\} = \int_{a'}^{b'} p(x)dx$$
- e isso significa precisamente que a variável aleatória ξ, que é a raiz da equação 5, tem uma função densidade de probabilidade $p(x)$

Geração de uma variável aleatória contínua

- **Exemplo:** Uma variável aleatória η será considerada *uniformemente distribuída no intervalo (a, b)* se sua densidade no intervalo for constante:

$$p(x) = 1/(b-a) \ para \ a < x < b.$$

- Para sortear o valor de η aplica-se a equação 5:

$$\int_a^\eta \frac{dx}{b-a} = \gamma$$

- Resolvendo-se a integral:

$$\frac{\eta - a}{b-a} = \gamma$$

- então:

$$\eta = a + \gamma(b-a) \tag{6}$$

Geração de uma variável aleatória contínua: método da rejeição

- Frequentemente a equação 5 é de difícil solução para ξ. Nesse caso convém encontrar um método que contorne a resolução direta dessa equação.

- Para isso foi desenvolvido o método de Von Neumann ou *método da rejeição*

Geração de uma variável aleatória contínua: método da rejeição

- Seja a função (complicada) que nos interessa f. Precisamos de uma outra função (facilmente integrável) g, tal que $g(x) > f(x)$ para qualquer x dentro do intervalo de interesse.
- Uma variável aleatória com probabilidade f pode ser gerada do seguinte modo:
  1. Sorteie um número aleatório X_i usando g como FDP
  2. Sorteie um número aleatório U_i distribuído uniformemente entre 0 e 1
  3. Se $f(X_i) \geq U_i * g(X_i)$, X_i é aceito
  4. Caso contrário X_i é rejeitado.
  5. De volta a (1) até que o númeo de pontos aceitos seja satisfatório.

Geração de uma distribuição normal.

- $$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\zeta} e^{-\frac{x^2}{2}}dx = \gamma$$

- Dada uma tabela de valores de ζ calculados anteriormente (*look-up table*) é simples demonstrar que:
$$\zeta' = a + \sigma\zeta \tag{7}$$
será também normal, com média $= a$ e variância $= \sigma^2$

W&J Cap 3: Estatísticas e Expectativas

- Nesse tópico estaremos embarcando num vasto mundo desenvolvido para a distribuição Normal em específico.
- Esse é o território clássico e, historicamente, foi desenvolvido pela baixa popularidade dos métodos bayesianos até a segunda metade do século XX.
- O uso de estatísticas não é particularmente fácil. Seus métodos são sutis, nada óbvios e frequentemente associados com uma pesada “maquinaria matemática”.
- Essa maquinaria será ignorada aqui, onde serão apresentados apenas seus resultados, ainda que trate-se de tornar claros os fundamentos conceituais subjacentes.

Estatísticas

- Conforme já visto anteriormente estatísticas são funções dos dados apenas que servem para sumariza-los, reduzi-los ou descreve-los.
- Estatísticas são, então, combinações de quantidades finitas de dados.
- A seguir valores fixos de dados em particular e funções dos dados apenas serão designados por letras maiúsculas (excepto para letras gregas). Valores possíveis e variáveis em letras minúsculas.

Estatísticas

- Podemos exemplificar estatísticas pelas que descrevem (a) a posição e (b) o espalhamento de um conjunto de dados.

- **Posição**
  - *Média:* $\overline{X} = 1/N \sum_{i=1}^{N} X_i$
  - *Mediana:* ordene X_i pelo seus valor e re-numere. então $X_{med} = X_j$, onde $j = N/2 + 0.5$ se N for ímpar e $X_{med} = (X_j + X_{j+1})/2$, onde $j = N/2$ se N for par.
  - *Moda:* X_{moda} é o valor de X_i que ocorre mais frequentemente; é a posição do pico do histograma de X_i

- **Espalhamento**
  - *Desvio médio:* $\overline{\Delta X} = 1/N \sum_{i=1}^{N} |X_i - X_{med}|$
  - *Desvio quadrático médio:* $S^2 = 1/N \sum_{i=1}^{N} (X_i - \bar{X})^2$
  - *Root-mean square (rms):* S

- Convém aqui lembrar que estatísticas realmente úteis são muitas raras. Para vários problemas de interesse estatísticas novas devem ser derivadas e o melhor modo para isso é pela inferência bayesiana.

Estatísticas

- Convém repetir que estatísticas são propriedades dos dados e não do modelo. Parâmetros como μ e σ das distribuições de Poisson e Normal definem essas distribuições e não estatísticas.
- De todos modo podemos saber de antemão que os nossos dados seguem uma dada distribuição e gostaríamos de relacionar as estatísticas com os parâmetros que descrevem essas distribuições.
- Isso é feito através dos *Valores Esperados* ou *Expectativas*
- O valor esperados $E[f(x)]$ de uma dada função f de uma variável aleatória x, com uma FDP g é definida como:

  $$E[f(x)] = \int f(x)g(x)\,dx,$$

  ou seja, a soma de todos os possíveis valores de f, ponderados pela sua probabilidade de ocorrência g.
- Podemos pensar no valor esperado como o resultado da média de um experimento repetidos muitas vezes. De fato, no caso de grandes números, a média $\bar{X}$ irá convergir para a média verdadeira, que é a valor esperado da função $f(x) = x$:

  $$E[x] = \int x g(x)\,dx$$

Estatísticas

- A estatística S^2 introduzida anteriormente deve convergir, do mesmo modo à variância, definida por:

$$E[(x-\mu)^2] = \int (x-\mu)^2 g(x)\, dx$$

- Outros valores esperados de grande importância são *enézimos momentos centrais*:

$$\mu_n = \int (x-\mu)^n g(x)\, dx$$

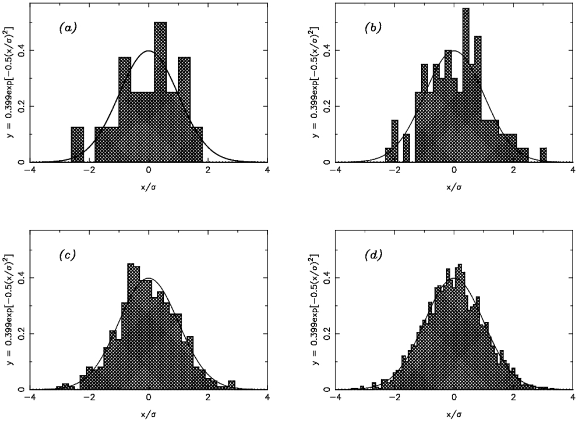
onde g é uma distribuição probabilística.

- Esses momentos são úteis para caracterizar a forma das distribuições, embora sejam muito suscetíveis a *intrusos*.

- Nesse contexto, as duas estatísticas mais usadas são a *skewness*, $\beta_1 = \mu_3^2$, que indica desvios da simetria e a *kurtosis*, $\beta_2 = \mu_3/\mu_2$, que indica a proeminência do pico (=3 para uma gaussiana).

O que esperar das estatísticas?

- Dado que frequentemente o número de dados é baixo que características as estatísticas devem ter para descreve-los do modo mais preciso possível, e mais eficientemente?



x_i gerados aleatoriamente a partir de uma distribuição Normal com $\mu = 0$ e $\sigma = 1$: (a) 20 valores, (b) 100 valores, (c) 500 valores e (d) 2500 valores. Os valores médios de x_i são, respectivamente, 0,003, 0,080, -0,032 e -0,005; os valores medianos 0,121, 0,058, -0,069 e -0,003; e os *rms* 0,968, 1.017, 0,986 e 1,001.

O que esperar das estatísticas?

1. Estatísticas devem ser *justas (unbiased)*, o que significa que o seu valor esperado deve ser o valor verdadeiro. Para X_i dados com uma distribuição gaussiana $\bar{X}$ é de fato uma estimativa justa da média, mas a estimativa justa da variância σ^2 é

$$\sigma_s = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2,$$

que difere do valor esperado de S^2 pelo fator $N/(N-1)$. Claramente, para $N \to \infty$ a diferença desaparece.

2. Deve ser *consistente*, que é o caso quando o valor é o verdadeiro para um grande número de dados. O *rms* pode não ser "justo", mas é consistente.

3. Deve obedecer a *proximidade*, isto é dar o o resultado mais próximo o possível da verdade. Por exemplo, a média não respeita esse critério para distribuição de Cauchy.

4. Deve ser *robusta*, ou seja resistente à presença de dados espúrios (intrusos ou *outliers*) nas caudas das distribuições. Falando sobre a posição a mediana é um estimador muito mais robusto do que a média.

5. Fim da aula 3