Probabilidades e processos. Noções básicas

R. Vilela Mendes
http://label2.ist.utl.pt/vilela/

Fevereiro 2010

Programa geral

1. Probabilidades e processos. Noções básicas

Programa geral

1. Probabilidades e processos. Noções básicas
2. O movimento Browniano

Programa geral

1. Probabilidades e processos. Noções básicas
2. O movimento Browniano
3. Introdução à matemática financeira

Programa geral

1. Probabilidades e processos. Noções básicas
2. O movimento Browniano
3. Introdução à matemática financeira
4. Nem tudo é Browniano, nem tudo é Gaussiano

Programa geral

1. Probabilidades e processos. Noções básicas
2. O movimento Browniano
3. Introdução à matemática financeira
4. Nem tudo é Browniano, nem tudo é Gaussiano
5. http://label2.ist.utl.pt/vilela/

Probabilidades e processos. Sumário

- Introdução
- Cadeias de Markov
- Variáveis Gaussianas
- Teorema do limite central
- Martingales
- Processos de ramificação
- Informação, entropia e medida de Gibbs
- Exercícios

Introdução

- $\boxed{\textbf{Modelos estocásticos}}$ **:** Descrição sistemática da dinâmica das variáveis desconhecidas
- Processos $\boxed{\textbf{discretos}}$ e $\boxed{\textbf{contínuos}}$
- Equações $\boxed{\textbf{progressivas}}$ e $\boxed{\textbf{regressivas}}$
  Relacionando, por exemplo, os valor expectáveis $f(x,t)$ dum porfólio obtido através de várias estratégias (equações matriciais, relações de recorrência, equações às derivadas parciais)
- Difusões e $\boxed{\textbf{cálculo de Ito}}$**:**
  Se $X(t)$ fôr diferenciável $\Delta X = X(t+\Delta t) - X(t)$ é de ordem Δt (existe C tal que $|X(t+\Delta t - X(t)| \leq C\,|\Delta t|$ para Δt pequeno), isto é $\Delta f(X(t)) = f(X(t+\Delta t)) - f(X(t)) \approx f'\Delta X$ to this accuracy.
- **Porém no cálculo de Ito ΔX é de ordem** $\sqrt{\Delta t}$, logo $\Delta f \approx f'\Delta X + \frac{1}{2}f''\Delta X^2$.

Introdução

- $\boxed{\text{Espaço de probabilidade } \Omega}$ $\omega \in \Omega$, $\omega =$ acontecimento elementar.
  $A \subset \Omega$, acontecimento
  $X(\omega) =$ variável aleatória
- **Probabilidade** de ω: $P(\omega) \geq 0$ com $\sum_{\omega \in \Omega} P(\omega) = 1$.
  $P(\omega)$ fracção de acontecimentos ω num grande número de ensaios
  $P(\omega)$ conhecimento subjectivo da ocorrência de ω

$$P(A) = \sum_{\omega \in A} P(\omega)$$

- Espaços de probabilidade $\boxed{\text{contáveis}}$ e $\boxed{\text{incontáveis}}$
  Se a probabilidade em $[0, 1]$ fôr uniforme, então $P(\omega) = 0$ para cada $\omega \in [0, 1]$.
- $\boxed{\text{Operações de conjunto}}$: Acontecimentos são subconjuntos de Ω, $A \subset \Omega$. Operações de conjunto aplicam-se aos acontecimentos
  $A \cap B, \quad A \cup B, \quad A \subseteq B, \quad A^c, \quad \emptyset.$
  $P(\emptyset) = 0$. O **complemento** de $\emptyset$ é Ω.
  A e B são **disjuntos** se $A \cap B = \emptyset$.

Introdução

- **Factos básicos:**
  $P(A) \leq P(B)$ se $A \subseteq B$,
  $P(A \cup B) = P(A) + P(B)$ se $P(A \cap B) = 0$,
  Se $P\,(\omega \neq 0)$ para todo o $\omega \in \Omega$, então $P(A \cap B) = 0$ se e só se A e B forem disjuntos,
  $P(A) + P(A^c) = P(\Omega) = 1$.
- $\boxed{\text{Probabilidade condicional}}$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- $\boxed{\text{Regra (ou teorema) de Bayes}}$: Permite calcular $P(A \cap B)$ quando forem conhecidos $P(B)$ e $P(A \mid B)$,

$$P(A \cap B) = P(A \mid B)P(B)$$

- **Independência:** A e B independentes se $P(A \mid B) = P(A)$. Isto é

$$P(A \cap B) = P(A) \cdot P(B)$$

Condicionar em B e depois em C é o mesmo que condicionar em $B \cap C$ (B e C)

$$
\begin{aligned}
P(\omega \mid B \mid C) &= \frac{P(\omega \mid B \cap C)}{P(C \mid B)} \\
&= \frac{P(\omega \cap C \mid B)}{\frac{P(C \cap B)}{P(B)}} \\
&= \frac{P(\omega \cap C \cap B)}{P(C \cap B)} \\
&= P(\omega \mid C \cap B)
\end{aligned}
$$

- No espaço de todos os acontecimentos possíveis haverá alguns em relação aos quais conhecemos a sua probabilidade e outros para os quais isso não é possível. Essa **informação incompleta** formaliza-se definindo a $\boxed{\text{algebra } F \text{ dos conjuntos}}$ mensuráveis que deve satisfazer as seguintes condições:
  $A \in F$ implica $A^c \in F$.
  $A \in F$ e $B \in F$ implica que $A \cup B \in F$ e $A \cap B \in F$.
  $\Omega \in F$ e $\emptyset \in F$.
- **Exemplo de** F: Suponhamos que se lança uma moeda 4 vezes mas que F se refere apenas aos dois primeiros lançamentos. Um acontecimento em F seria $\{00\} = \{0000, 0001, 0010, 0011\}$
  Um acontecimento não pertencente a F seria, por exemplo
  $A = \{1111, 1110, 1101, 1011, 0111\}$
- $\boxed{\textbf{Algebra}-\sigma}$ = fechada para intersecções contáveis: Se $A_n \in F$ é um conjunto contável mensurável em F então $A = \cap_n A_n \in F$. Uma algebra finita é sempre σ. Se Ω for infinito pode ou não ser σ.

Introdução

- Exemplo: $\Omega =$ conjunto dos inteiros e $A \in F$ se A é finito ou A^c é finito. F é uma algebra que não é σ. Por exemplo, se A_n exclui só os primeiros n inteiros impares, então $A = \cap_n A_n$ é o conjunto dos inteiros pares e nem A nem A^c são finitos.
- $\boxed{\text{Função de distribuição cumulativa}}$ $F(x)$, é a probabilidade que $X \leq x$,

$$F(x) = \sum_{X(\omega) \leq x} P(\omega)$$

- $\boxed{\text{Funções mensuráveis}}$: $X(\omega)$ é mensurável em relação a F se o valor de X é completamente determinada pela informação em F. Isto é, para cada valor x, o acontecimento $X = x$, $B_x = \{\omega \ : \ X(\omega) = x\} \in F$. A função $X(\omega)$ é mensurável se todos os conjuntos B_x são mensuráveis. Pode escrever-se $X \in F$.
- $\boxed{\text{Algebra gerada por acontecimentos}}$ $A_1, \ldots, A_k$: a menor algebra que contém estes acontecimentos, todas as intersecções, uniões e complementos.

Introdução

- $\boxed{\text{Algebra } F_X \text{ gerada por uma função}}$: algebra gerada pelos conjuntos $B_x = \{\omega \ : \ X(\omega) = x\}$.
  $Y \in F_X$ se o conhecimento de $X(\omega)$ determina o valor de $Y(\omega)$.
- **Relação de equivalência**: ω_1 e ω_2 são equivalentes, $\omega_1 \sim \omega_2$, se a informação em F não distingue ω_1 de ω_2. $\omega_1 \sim \omega_2$ se $\omega_1 \in A \Rightarrow \omega_2 \in A$ para todo o $A \in F$.
- **Partição**: Uma partição de Ω é $P = \{B_1, B_2, \dots\}$ tal que todo o $\omega \in \Omega$ está exactamente num dos acontecimentos B_k. $F_P =$ algebra$-\sigma$ gerada por P.
- **Valor expectável** (valor médio) de $X(\omega)$:
  $E[X] = \sum_{\omega \in \Omega} X(\omega) P(\omega)$
  $E[X_1 + X_2] = E[X_1] + E[X_2]$ e $E[cX] = cE[X]$.
- $\boxed{\text{O valor expectável é a melhor aproximação quadrática}}$
  $E\left[(X-x)^2\right] = E[X^2 - 2Xx + x^2] = E[X^2] - 2xE[X] + x^2$ e minimizando obtém-se $x_{opt} = E[X]$

- **Valor expectável condicional**:

$$E[X|B] = \sum_{\omega \in B} X(\omega) P(\omega|B)$$

- **A lei da probabilidade total**: $P = \{B_1, B_2, \ldots\}$ partição de Ω,

$$\boxed{E[X] = \textstyle\sum_k E[X \mid B_k] P(B_k)}$$

$$\begin{aligned}
E[X] &= \sum_{\omega \in \Omega} X(\omega) P(\omega) \\
&= \sum_k \left(\sum_{\omega \in B_k} X(\omega) P(\omega) \right) \\
&= \sum_k \left(\sum_{\omega \in B_k} X(\omega) \frac{P(\omega)}{P(B_k)} \right) P(B_k) \\
&= \sum_k E[X \mid B_k] P(B_k) \, .
\end{aligned}$$

Introdução

Variância:

$$Var\left(X\right)=E\left[\left(X-E\left[X\right]\right)^{2}\right]=E\left[X^{2}\right]-\left(E\left[X\right]\right)^{2}$$

Covariância de duas variáveis aleatórias:

$$cov\left(X,Y\right)=E\left[\left(X-E\left[X\right]\right)\left(Y-E\left[Y\right]\right)\right]=E\left[XY\right]-E\left[X\right]E\left[Y\right]$$

Momento de ordem n:

$$m_n=E\left[X^n\right]$$

Função característica:

$$\phi_X\left(s\right)=E\left[e^{isX}\right]$$

(Função geradora dos momentos $E\left[e^{\lambda X}\right]$)
Momentos e função característica:

$$m_n=\frac{1}{i^n}\frac{d^n}{ds^n}\left.\phi_X\left(s\right)\right|_{s=0}$$

Desigualdades:
Chebyshev: Se $\lambda > 0$

$$P(|X| > \lambda) \leq \frac{1}{\lambda^p} E[|X|^p]$$

Schwartz:

$$E[XY] \leq \sqrt{E[X^2]\, E[Y^2]}$$

Hölder:

$$E[XY] \leq (E[|X|^p])^{\frac{1}{p}} (E[|X|^q])^{\frac{1}{q}}$$

$p, q > 1$ e $\frac{1}{p} + \frac{1}{q} = 1$
Jensen: Se $\phi : R \rightarrow R$ fôr uma função convexa e X e $\phi(X)$ tiverem valor expectável finito

$$\phi(E[X]) \leq E[\phi(X)]$$

Convergências

- **Convergência quase segura:** $X_n \overset{q.s.}{\longrightarrow} X$, se

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

para todo o $\omega \notin N$, em que $P(N) = 0$

- **Convergência em probabilidade:** $X_n \overset{P}{\longrightarrow} X$, se

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

para todo o $\varepsilon > 0$

- **Convergência em média de ordem** $p \geq 1$ **:** $X_n \overset{L^p}{\longrightarrow} X$, se

$$\lim_{n\to\infty} E\left[|X_n - X|^p\right] = 0$$

- **Convergência em lei:** $X_n \overset{\mathcal{L}}{\longrightarrow} X$, se

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

para todos os pontos x em que a função de distribuição F fôr contínua.

Introdução

- A convergência em média de ordem p implica convergência em probabilidade.
- A convergência quase segura implica convergência em probabilidade
- Convergência em probabilidade implica a existência de uma subsequência que tem convergência quase segura
- Convergência quase segura implica convergência em média de ordem $p \geq 1$ se as variáveis aleatórias X_n forem limitadas em valor absoluto por uma variável aleatória Y com momento de ordem p finito.
- Convergência em probabilidade implica convergência em lei.

- Um **processo estocástico** é uma colecção de variáveis aleatórias $\{X_t, t \in T\}$ definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{F}, P)$. Uma cadeia de Markov é um dos processos estocásticos mais simples.
- $\boxed{\text{Estados e probabilidades de transição entre estados}}$
  A probabilidade de transição depende apenas do estado em que se está. Tempo discreto, $0, \cdots, t, t+1, t+2, \cdots$.
  $S =$ espaço de estados $(s_1, \ldots, s_m,$ ou simplesmente $1, 2, \ldots, m)$
  $X(t)$ ou $X_t =$ estado no tempo t
  $X = (X_0, X_1, \ldots, X_T) = (X(0), X(1), \ldots, X(T)) =$ trajectória. As trajectórias são os acontecimentos ω e o espaço de probabilidade Ω é o espaço das trajectórias.
- $\boxed{\text{Algebras } F_t}$: Informação colhida pela observação da cadeia de Markov até ao tempo t.
  $\boxed{\text{Algebra } G_t}$: algebra gerada por $X(t)$ apenas
- X_1 and X_2 equivalentes em F_t se $X_1(s) = X_2(s)$ para $0 \leq s \leq t$.
  $F_t \subseteq F_{t+1}$, $G_t \subseteq F_t$, mas $G_t \nsubseteq G_{t+1}$, $(X(t+1)$ não determina $X(t))$

- $\boxed{\text{Funções adaptadas ou não-anticipatórias}}$
  $F(X,t)$ é adaptada se para cada t ela é mensurável em relação a F_t $=$ é determinada pelos valores de $X(s)$ para $s \leq t$.
- **Exemplos**:
  (1) $F_1(X,t) = 1$ se $X(s) = 1$ para $s \leq t$, $F_1(X,t) = 0$ caso contrário
  (2) $F_2(X,t) = \min(s > t)$ em que $X(s) = 1$ ou $F_2(X,t) = T$ se $X(s) \neq 1$ para $t < s \leq T$.
  F_1 é adaptada a F_t, F_2 não é.
- Funções adaptadas são modelos de decisão com informação incompleta. Uma função F ser adaptada significa que a decisão se faz baseada na informação disponível até ao tempo t.
- $\boxed{\text{Propriedade Markoviana}}$: $X(t)$ é toda a informação que é util para prever o futuro,
  $P(X_{t+1} = k | X_t = j) = P(X_{t+1} = k | X_t = j, X_{t-1} = l, etc.)$
  ou, $P(X_{t+1} = k \mid \mathcal{F}_t) = P(X_{t+1} = k \mid \mathcal{G}_t)$

Cadeias de Markov

- $\boxed{\text{Probabilidade de transição}}$

$$P_{jk} = P\left(X(t+1) = k \mid X(t) = j\right) = P(j \to k)$$

- Cadeia de Markov **estacionária** se os P_{jk} não dependem de t.
  $\sum_k P_{jk} = 1$
- $\boxed{\text{Probabilidade das trajectórias}}$ Seja $f_0(j) = P(X(0) = j)$.

$$\begin{aligned} P(X(1) = k, X(0) = j) \\ = \quad P(X(1) = k \mid X(0) = j) \cdot P(X(0) = j) = f_0(j) P_{jk} \ . \end{aligned}$$

Para dois passos no tempo:

$$\begin{aligned} P(X(2) = l, X(1) = k, X(0) = j) \\ = \quad & P(X(2) = l \mid X(1) = k, X(0) = j) \cdot P(X(1) = k, X(0) = j) \\ = \quad & P(X(2) = l \mid X(1) = k) \cdot P(X(1) = k, X(0) = j) \\ = \quad & f_0(j) P_{jk} P_{kl} \ . \end{aligned}$$

Probabilidades das trajectórias obtidas por multiplicação das $\{P_{jk}\}$

Cadeias de Markov

- $\boxed{\text{Matrix de transição}}$: $P = \{P_{jk} = P(j \to k)\}$
  $\sum_k P_{jk} = 1$.
- Uma matrix em que os elementos são não-negativos e as linhas somam 1, diz-se uma $\boxed{\text{matrix estocástica}}$

$$P_{jk}^s = P(X_{t+s} = k \mid X_t = j)$$

- P_{jk}^s é o elemento (j, k) da potência s (P^s) de P
- $\boxed{\text{Equações progressivas e regressivas}}$ para as cadeias de Markov
  Seja $u(k, t) = P(X(t) = k)$ e
  $f(k, t) = E[V(X(T)) \mid X(t) = k]$ (para $t < T$)
  Equações progressivas: Cálculo de $u(k, t+1)$ quando os $u(j, t)$ são conhecidos (do tempo t para o tempo $t+1$)
  Equações regressivas: Cálculo de $f(k, t)$ quando os $f(j, t+1)$ são conhecidos (do tempo $t+1$ para o tempo t)

- $\boxed{\text{Equação progressiva}}$:

$$\begin{aligned} u(k, t+1) &= P(X(t+1) = k) \\ &= \sum_j P(X(t+1) = k \mid X(t) = j) \cdot P(X(t) = j) \ . \end{aligned}$$

donde

$$u(k, t+1) = \sum_j P_{jk} u(j, t)$$

Também se chama a **equação de Chapman-Kolmogorov**.

Em forma matricial, sendo $u\left(t+1\right)$ e $u\left(t\right)$ vectores linha

$$u(t+1) = u(t)P$$

Por iteração da multiplicação de matrizes

$$u(t) = u(t-1)P = u(t-2)P^2 = \cdots = u(0)P^t$$

- $\boxed{\text{Equação regressiva}}$:
  Perspectiva de obter $V(X_T)$ no fim dum periodo dado o que está a suceder em t, $E\left[V(X_T) \mid X_t = k\right]$
  Para $t \leq T$, defina-se

$$f(k,t) = E_{k,t}[V(X_T)] = E\left[V(X_T) \mid X_t = k\right]$$

$$\begin{aligned} f(k,t) &= E\left[V(X_T) \mid X_t = k\right] \\ &= \sum_{j\in\mathcal{S}} E\left[V(X_T) \mid X_t = k, X_{t+1} = j\right] \cdot P(X_{t+1} = j \mid X_t = k) \\ f(k,t) &= \sum_{j\in\mathcal{S}} f(j, t+1) P_{kj} \ . \end{aligned} \tag{1}$$

O último passo utiliza a propriedade Markoviana

$$E[V(X_T) \mid X_t = k, X_{t+1} = j] = E_{j,t+1}[V(X_T)] \ .$$

A dinâmica (1) em conjunto com a condição final

$$f(k,T) = V(k)$$

permite determinar $f(k,T-1)$, e sucessivamente $f(k,T-2)$, etc.

- $\boxed{\text{Versão matricial}}$:
  $f(k,t)$ vector coluna, $f(t) = (f(1,t), f(2,t), \cdots)^t$

  $$f(t) = Pf(t+1)\ .$$

  e usando a propriedade associativa do producto de matrizes

  $$f(t) = P^{T-t}V$$

  V é o vector coluna dos $V(k)$

  $$\begin{aligned} E[V(X_T)] &= \sum_{k\in\mathcal{S}} P(X_t = k)\cdot E[V(X_T)\mid X_t = k] \\ &= \sum_{k\in\mathcal{S}} u(k,t)f(k,t) \\ &= u(t)f(t)\ . \end{aligned}$$

Variável Gaussiana escalar: É uma variável aleatória X com densidade de probabilidade

$$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

com factor de normalização $\frac{1}{\sqrt{2\pi}}$ porque $\int_{-\infty}^{\infty} \rho(x)dx = 1$

$$E[X] = 0$$

Variância:

$$\begin{aligned} E[X^2] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left(x e^{-x^2/2} \right) dx \\ &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left(\frac{d}{dx} e^{-x^2/2} \right) dx \\ &= -\frac{1}{\sqrt{2\pi}} \left(x e^{-x^2/2} \right) \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1 \end{aligned}$$

Também: $E[X^4] = 3$, $E[X^6] = 15$,

Uma variável Gaussiana com média $E[X] = \mu$ e variância $E(X) = E[(X-\mu)^2] = \sigma^2$ tem densidade

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ .$$

que habitualmente se representa por $X \sim N(\mu, \sigma^2)$

Variável Gaussiana vectorial:

$$\rho(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det\left(H^{-1}\right)}} e^{-\frac{1}{2}x^* H x} \ ,$$

Sendo H uma matrix $n \times n$, simétrica e positiva $\left(x^* H x > 0\right)$ e x um vector com n dimensões,

No caso de ter média μ, $x^* H x$ é substituido por $(x-\mu)^* H (x-\mu)$.

Função característica:

$$\begin{aligned} C_n(\xi) &= E\left[e^{ix \bullet \xi}\right] = \int_{-\infty}^{\infty} e^{ix \bullet \xi} \rho(x)\, d^n x \\ C_1(\xi) &= e^{-\frac{\xi^2}{2}} \qquad\qquad C_n(\xi) = e^{-\frac{1}{2}\xi^* H^{-1} \xi} \end{aligned}$$

O teorema do limite central

A importância das variáveis Gaussianas é uma consequência do seguinte **Teorema:** Sejam $X_1, X_2, \cdots, X_n, \cdots$ variáveis aleatórias independentes com a mesma função de distribuição, com média μ e variância $\sigma^2 < \infty$. Então se $S_n = X_1 + X_2 + \cdots + X_n$

$$P\left(\frac{S_n - n\mu}{\sqrt{n}} \leq x\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\rightarrow}} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{\sigma^2}}$$

Prova: Faça-se, sem perda de generalidade $\mu = 0$ $(X_i \to X_i - \mu)$. Como as variáveis são independentes a função característica da soma é o producto das funções características

$$E\left[e^{i\lambda\frac{S_n}{\sqrt{n}}}\right] = \left(C\left(\frac{\lambda}{\sqrt{n}}\right)\right)^n$$

Usando

$$C\left(\frac{\lambda}{\sqrt{n}}\right) = E\left[e^{i\frac{\lambda}{\sqrt{n}}x}\right] = 1 + i\frac{\lambda}{\sqrt{n}}E\left[x\right] - \frac{\lambda^2}{2n}E\left[x^2\right] - i\frac{\lambda^3}{3!n\sqrt{n}}E\left[x^3\right] + \cdots$$

O teorema do limite central

$$\left(C\left(\frac{\lambda}{\sqrt{n}}\right)\right)^{n} \to \left(1-\frac{\lambda^{2}\sigma^{2}}{2n}+o\left(\frac{1}{n}\right)\right)^{n} \to e^{-\frac{1}{2}\sigma^{2}\lambda^{2}}$$

Note-se o papel importante desempenhado pelo facto de $E\left[x^{2}\right]$ ser finito. No caso mais geral em que as variâncias podem ser diferentes, mas finitas, seja

$$B_n^2 = \sum_{k=1}^{n} \sigma_k^2$$

então a distribuição de $\left(\frac{X_1+X_2+\cdots+X_n}{B_n}\right)$ converge para uma distribuição Gaussiana com $\sigma^2 = 1$

Componentes principais e independência

Dadas duas variáveis aleatórias X_1 e X_2 com valores expectáveis

$$\overline{X_1} = E\left[X_1\right] \qquad\qquad \overline{X_2} = E\left[X_2\right]$$

calculemos

$$M_{12} = E\left[\left(X_1 - \overline{X_1}\right)\left(X_2 - \overline{X_2}\right)\right]$$

Se esta quantidade (a covariância das variáveis centradas) fôr nula, diz-se que as variáveis estão **descorrelacionadas.**

Variáveis independentes são descorrelacionadas mas o contrário não é necessáriamente verdade. A excepção são as variáveis Gaussianas. Se a covariância for nula a matrix H^{-1} em

$$e^{-\frac{1}{2}x^{*}Hx}$$

é diagonal e a probabilidade é o produto das probabilidades, por causa da propriedade de factorização da exponencial.

Componentes principais e independência

Componentes principais: Dada uma matrix de correlação M dum conjunto de variáveis aleatórias, diagonaliza-se e determinam-se os vectores próprios

$$Mv_j = \lambda_j v_j$$

os v_j's são combinações lineares das variáveis originais. Por força da diagonalização a matrix de covariância dos v_j's tem elementos não-diagonais nulos. Portanto os v_j's são variáveis descorrelacionadas. A esta método de obtenção de componentes não-correlacionadas num conjunto de variáveis chama-se **análise em componentes principais** (PCA). Se as variáveis não forem Gaussianas isto não garante a independência das componentes. Há então que usar outros métodos para obter as componentes independentes (ICA).

A algebra $\mathcal{F}_t$ representa toda a informação disponível (*todos os acontecimentos sobre os quais se pode ter alguma informação*) até ao tempo t. Supondo que ao passar do tempo t para o tempo $t+1$ não se perdeu informação sobre o passado e possivelmente se acrescentou alguma, teremos

$$\mathcal{F}_t \subset \mathcal{F}_{t+1}$$

e assim sucessivamente. A uma sequência de algebras que satisfaz esta propriedade chama-se uma **filtragem**.

Martingale: Um processo M (adaptado a $\mathcal{F}_t$) diz-se uma martingale se

$$E\left[M_{t+1} | \mathcal{F}_t\right] = M_t$$

Usando o facto que condicionamentos sucessivos são equivalentes ao condicionamento em relação à intersecção

$$E\left[M_{t+2} | \mathcal{F}_t\right] = E\left[\left(E\left[M_{t+2} | \mathcal{F}_{t+1}\right]\right) | \mathcal{F}_t\right] = E\left[M_{t+1} | \mathcal{F}_t\right] = M_t$$

conclui-se que para todo o $s > 0$ $\qquad E\left[M_{t+s} | \mathcal{F}_t\right] = M_t$

Martingales

As **inovações** ($Y_{t+1} = M_{t+1} - M_t$) duma martingale têm valor expectável zero

$$E\left[Y_{t+1} | \mathcal{F}_t\right] = 0$$

Se o processo $U(\omega) \in \mathcal{F}$

$$E\left[UM | \mathcal{F}\right] = UE\left[M | \mathcal{F}\right]$$

Consideremos o processo

$$G_t = U_0 Y_1 + U_1 Y_2 + \cdots + U_{t-1} Y_t$$

em que $U \in \{\mathcal{F}_t\}$. Se Y_{t+1} fôr a diferença de preço duma acção entre o tempo t e o tempo $t+1$, G_t poderia ser o ganho acumulado no mercado quando no intervalo (t_s, t_{s+1}) se detêm U_s unidades dessa acção. Pela propriedade anterior se M_t fôr uma martingale

$$E\left[U_t Y_{t+1} | \mathcal{F}_t\right] = 0$$

e G_t também é uma martingale.

A noção de mercado eficiente:
Por exemplo: $\{\mathcal{F}_t\}$ = informação sobre a história dos preços das acções no mercado até ao tempo t
$\{\mathcal{G}_t\}$ = a informação anterior mais todos os dados possíveis sobre o estado da economia até ao tempo t
$\{\mathcal{F}_t\} \subset \{\mathcal{G}_t\}$. Um processo pode ser uma martingale em relação a $\{\mathcal{F}_t\}$ mas não ser em relação a $\{\mathcal{G}_t\}$
Mercado eficiente (noção fraca) $e^{-\mu t} S(t)$ é uma $\{\mathcal{F}_t\}$ $-$martingale
Mercado eficiente (noção forte) $e^{-\mu t} S(t)$ é uma $\{\mathcal{G}_t\}$ $-$martingale
(μ é a taxa média de crescimento da economia $\backsim$ juro dum depósito a prazo ou duma obrigação do tesouro)

Processo de Galton-Watson

$p_k(t)$ = probabilidade de existência de k indivíduos no tempo t

Estuda-se a proliferação duma família que descende dum único indivíduo

Condição inicial: $p_1(0) = 1$ e $p_k(0) = 0$ se $k \neq 1$. A probabilidade da família se extinguir no tempo t é $p_0(t)$.

Seja λ a taxa de nascimento e μ a taxa de morte. Então

$$\frac{d}{dt}p_0 = \mu p_1$$

e para $k \neq 0$

$$\frac{d}{dt}p_k = \lambda (k-1) p_{k-1} - (\lambda + \mu) k p_k + \mu (k+1) p_{k+1}$$

Sendo $\langle k \rangle = \sum_{k=0}^{\infty} k p_k$ obtém-se

$$\frac{d}{dt}\langle k \rangle = (\lambda - \mu) \langle k \rangle$$

$(\lambda - \mu)$ é taxa esperada de crescimento do valor médio da população.

Fluctuações:

$$v = \frac{\sqrt{\langle k^2 \rangle - \langle k \rangle^2}}{\langle k \rangle}$$

$$\frac{d}{dt} \langle k^2 \rangle = 2(\lambda - \mu) \langle k^2 \rangle + (\lambda + \mu) \langle k \rangle$$

$\lambda = \mu$ é o ponto crítico do processo de ramificação.
Se $\gamma = \lambda - \mu = 0$ então $\langle k \rangle = 1$ e $\langle k^2 \rangle = 1 + 2\mu t \Longrightarrow v = \sqrt{2\mu t}$
Se $\gamma = \lambda - \mu > 0$ então $\langle k \rangle \backsim \exp(\gamma t) \to \infty$ e
$\langle k^2 \rangle \backsim \exp(-\gamma t/2) \to 0$
Se $\gamma = \lambda - \mu < 0$ então $\langle k \rangle \backsim \exp(\gamma t) \to 0$ e
$\langle k^2 \rangle \backsim \exp(-\gamma t/2) \to \infty$
Portanto se $\gamma = \lambda - \mu \leq 0$ o sistema é dominado pelas fluctuações.
Solução para o caso $\lambda = \mu$ $\quad p_0 = \frac{\mu t}{1+\mu t}$

$$p_k = \frac{(\mu t)^{k-1}}{(1 + \mu t)^{k+1}}$$

Processos de ramificação

Seja $\lambda = \mu$. Para tempos grandes $(\mu t >> 1)$ a maior parte das famílias extingue-se $(p_0 \to 1$ e $p_k \to 0)$. Isto é, para tempos grandes, o número mais provável de membros duma família é zero. Por outro lado para $\lambda = \mu$ o valor médio da população seria $\langle k \rangle = 1$
Suponhamos agora que temos no tempo zero N_0 indivíduos cada um deles dando origem a uma família. O número médio de famílias que sobrevivem é

$$\langle F \rangle = N_0 \left(1 - p_0\right) = \frac{N_0}{1 + \mu t}$$

Vejamos qual o comportamento duma família sobrevivente. Se sobrevive, a distribuição de probabilidade do número de indivíduos n numa familia sobrevivente será

$$p_n = \frac{(\mu t)^{n-1}}{(1 + \mu t)^{n+1}}$$

$n = 1, 2...$

Processos de ramificação

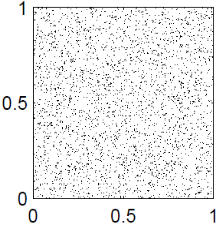
Donde

$$\begin{aligned} \langle n(t) \rangle &= \mu t + 1 \\ \langle n^2 \rangle &= (2\mu t + 1)(\mu t + 1) \end{aligned}$$

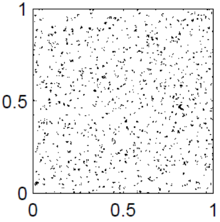
Portanto o número de famílias sobreviventes diminui com $\frac{1}{t}$ enquanto que o número de elementos duma família sobrevivente cresce com t.
Isto levou Galton e Watson à conclusão de que "*a diminuição observada no número de apelidos das famílias cuja história pode ser seguida, não é um sinal da diminuição da sua fertilidade*". Em vez disso a extinção da maioria das famílias e a aparente fecundidade de algumas, poucas, excepções é o "*resultado das leis do acaso*".
Este tipo de processo conjugado com a difusão espacial explica a emergência de concentrações e vazios na dinâmica de populações (ecologia, oceanografia, etc.)
(Na Coreia há cerca de 250 nomes de família e 45% da população corresponde a apenas 3 nomes de família)

Processos de ramificação: Simulação dum superprocesso

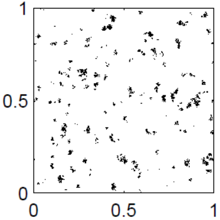
t=0



t=10



t=100



Processos de ramificação

Mr. Galton then read the following paper by the Rev. H. W. Watson and himself:

On the PROBABILITY *of the* EXTINCTION *of* FAMILIES. By the Rev. H. W. WATSON. With PREFATORY REMARKS, by FRANCIS GALTON, F.R.S.

THE decay of the families of men who occupied conspicuous positions in past times has been a subject of frequent remark, and has given rise to various conjectures. It is not only the families of men of genius or those of the aristocracy who tend to perish, but it is those of all with whom history deals, in any way, even of such men as the burgesses of towns, concerning whom Mr. Doubleday has inquired and written. The instances are very numerous in which surnames that were once common have since become scarce or have wholly disappeared. The tendency is universal, and, in explanation of it, the conclusion has been hastily drawn that a rise in physical comfort and intellectual capacity is necessarily accompanied by diminution in "fertility"—using that phrase in its widest sense and reckoning abstinence from marriage as sterility. If that conclusion be true, our population is chiefly maintained though the "proletariat," and thus a large element of degradation is inseparably connected with those other elements which tend to ameliorate the race. On the other hand, M. Alphonse De Candolle has directed attention to the fact that, by the ordinary law of chances, a large proportion of families are continually dying out, and it evidently follows that, until we know what that proportion is, we cannot estimate whether any observed diminution of surnames among the families whose history we can trace, is or is not a sign of their diminished "fertility." I give extracts from M. De Candolle's work in a foot-note,* and may add that, although I have not hitherto published anything on the matter, I took considerable pains some years ago to obtain numerical results in respect to this very problem. I made certain very simple, but not very inaccurate, suppositions, concerning average fertility, and I worked to the nearest integer, starting with 10,000 persons, but the computation became intolerably tedious after a few steps, and I had to abandon it. More recently, having first privately applied in vain to some

* "Au milieu des renseignements précis et des opinions très-sensées de MM. Benoiston de Châteauneuf, Galton, et autres statisticiens, je n'ai pas rencontré la réflexion bien importante qu'ils auraient dû faire de l'extinction *inévitable* des noms defamille. Évidemment tous les noms doivent s'éteindre ....Un mathématicien pourrait calculer comment la réduction des noms ou titres aurait lieu, d'après la probabilité des naissances toutes féminines

Processos de ramificação

Superprocesses and Plankton Dynamics

Robert Adler
Faculty of Industrial Engineering and Management, Technion, Haifa, Israel
Department of Statistics, University of North Carolina, Chapel Hill, North Carolina

Abstract. This paper is a first introduction to superprocesses for non-probabilists. They are presented so as to suggest natural models for the growth and motion of large plankton populations in either stable or flowing ocean environments.

Introduction

For the sake of academic honesty, I should start this paper with a disclaimer: Although the title contains both the keywords "superprocesses" and "plankton dynamics", and I do know something about the former, my knowledge of the latter, and of biological oceanography in general, starts and stops at the level of the elementary textbook Lalli and Parsons (1993). This would not generally be considered a level of expertise that would allow an author to start writing on a topic. Nevertheless, it does seem to me that there is a natural connection between the two topics, and that both are likely to benefit from interdisciplinary cross-fertilisation. This point was emphasised, for me, by a number of talks at the *Aha Huliko'a* meeting, and I shall return to this and some general philosophy on

tantly, how to place the entire superprocess structure within a randomly moving frame of reference such as an oceanographic flow. The final Section 6 has the promised general comments on stochastic modelling, motivated by the superprocess models of the remainder of the paper.

It does not seem to make too much sense in an introductory paper of this kind to worry too much about assigning detailed credits for the structures and results that will be described. Since the interested reader will have to go elsewhere for missing details, (s)he can search for references at the same time. Without doubt, the best place to start is with the Saint Flour lectures of Dawson (1993), which has a properly credit-assigned and almost encyclopædic treatment of superprocess accurate up to the time of its writing.

Informação, entropia e medida de Gibbs

Informação = Medida do conhecimento que se adquire com a ocorrência de um acontecimento quando previamente apenas se conhecia a sua probabilidade

Suponhamos N acontecimentos indexados por uma memória binária de comprimento b $\left(N = 2^b \qquad b = \log N / \log 2\right) \qquad b = \log_2 N$ é o número de bits.

Consideremos os acontecimentos agregados em classes $N = \sum_i N_i$

$p_i = N_i / N$ é a probabilidade da classe i.

Seja b_i a informação recebida quando nos dizem que o acontecimento pertence à classe i

Portanto $b_i + \log_2 N_i = \log_2 N$, isto é, a informação recebida mais o número de bits que ainda falta especificar deve ser igual à incerteza inicial. Daqui conclui-se

$$b_i = -\log(N_i / N) = -\log p_i$$

Portanto o valor médio da informação associada aos acontecimentos é

$$I(p) = \sum_i p_i \log p_i$$

$S = -I$ (incerteza = entropia). A

$$S(p) = \sum_i -p_i \log p_i$$

chama-se **Entropia de Shannon**.
A entropia é um instrumento poderoso para estimar probabilidades dum acontecimento quando apenas se tem informação incompleta.
Suponhamos, por exemplo, que se conhece apenas o valor médio $\langle X \rangle$ duma variável

$$\sum_i X_i p(X_i) = \langle X \rangle \tag{2}$$

Por outro lado sabe-se que

$$\sum_i p(X_i) = 1 \tag{3}$$

A melhor alternativa é incorporar a informação conhecida e minimizar a informação sobre aquilo que não se conhece, isto é, fazer uma previsão o menos preconceituosa possível (**Princípio da entropia máxima**)

Maximizar

$$\Gamma = \sum_i \{-p(X_i)\log p(X_i) - \lambda_1 p(X_i) - \lambda_2 X_i p(X_i)\}$$

Donde

$$\partial\Gamma/\partial p(X_i) = -\log p(X_i) - 1 - \lambda_1 - \lambda_2 X_i = 0$$

$$p(X_i) = e^{-\lambda_2 X_i} e^{-(\lambda_1 + 1)}$$

sendo λ_1 e λ_2 obtidos a partir de (2) e (3).
Uma vez que $\partial^2\Gamma/\partial p(X_i)^2 = -1/p(X_i) < 0$ é de facto um máximo.
A esta distribuição de probabilidade chama-se **medida de Gibbs**.

Exercícios

1 - Numa caixa existem quatro bilhetes em que está escrito, respectivamente, AAB, ABA, BAA, BBB. Escolhe-se um bilhete ao acaso. Considere os seguintes acontecimentos E1={A aparece em primeiro}, E2={A aparece em segundo}, E3={A aparece em terceiro lugar}. Mostre que estes acontecimentos são independentes dois a dois mas não são globalmente independentes.

2 - Uma doença ocorre em 2 de cada 10000 habitantes ($p_0(A) = 0,0002; p_0(B) = 0,9998$). Existe um teste com fiabilidade $99,5\%$, mas com 2 falsos positivos em cada 1000 casos saudáveis ($p(x|A) = 0,995; p(x|B) = 0,002$).
Se um indivíduo tiver um teste positivo, qual é a probabilidade de efectivamente ter a doença ?

Exercícios

3 - Admita-se que cerca de $1,5\%$ dos atletas de alta competição recorre a estimulantes $(p_0(A) = 0,015; p_0(B) = 0,985)$. Considere-se um tipo de análise que obtém resultados positivos em
99% dos casos e tem falsos positivos em 4% dos casos
$(p(x|A) = 0,99; p(x|B) = 0,04)$
a) Qual a probabilidade de um atleta com teste positivo se ter efectivamente drogado ?
b) No caso de se repetir o teste e ele dar de novo positivo qual é então a probabilidade de ele se ter drogado ?
$\left(p(xx|A) = p(x|A)^2; p(xx|B) = p(x|B)^2\right)$

4 - Há uma variável aleatória X que pode tomar os valores $(1,2,3,4,5,6)$. Esta variável é amostrada 100000 vezes e no final a média $\left(\frac{1}{100000}\sum_{i=1}^{100000} X_i\right)$ foi $3,83462$. O número de vezes que se obteve 2 foi 13742. Qual é a melhor estimativa para o número de vezes em que se obteve 1?

5 - Uma variável x pode tomar 4 posições diferentes $(1, 2, 3, 4)$ ao longo dum circulo (Exemplos: direcções de navegação dum submarino, direcções de deslocação dum cardume, posições dum camião de entrega de mercadorias, etc.). Em cada intervalo de tempo a probabilidade de se deslocar para a frente ou para trás é p e de ficar no mesmo lugar é q $(q + 2p = 1)$.
a) Calcule a probabilidade de se encontrar em 3 ao fim de dez unidades de tempo se o ponto inicial for 1.
b) Determine uma formula geral para essa probabilidade de transição em n unidades de tempo.
c) Determine o limite dessa probabilidade quando $n \to \infty$

6 - Em $t = 0$, um milhão de organismos é distribuido uniformemente numa área de um kilómetro quadrado. A sua taxas de morte e reprodução são iguais $(\lambda = \mu)$. Ao fim de um tempo T escolhe-se ao acaso uma área de um metro quadrado. Qual é a probabilidade de aí encontrar um organismo vivo?