...

Calcul de probabilité - Calcul d'intégrale

Probabilité à calculer sur X :

$$\mathbb{P}\left[X = x\right]$$

Souvent on connait bien la probabilité conditionnelle de X sachant un paramètre θ :

$$\mathbb{P}\left[X = x \mid \theta\right]$$

Par exemple, si X est le nombre de mutations accumulées par une séquence pendant un temps t :

$$\mathbb{P}\left[X = x \mid T = t\right] = \exp\left(-u\ t\right)\frac{(u\ t)^x}{x!}$$

loi de Poisson de paramètre $(u\ t)$ avec u un taux de mutation.

Calcul de probabilité - Calcul d'intégrale

Théorème des probabilités totales :

$$\mathbb{P}\left[X = x\right] = \int_t \mathbb{P}\left[X = x \mid T = t\right] \, f_T(t) \, dt$$

Le problème devient donc un problème de calcul d'intégrale.

Vraisemblance

La fonction de vraisemblance $L(\theta)$ d'un paramètre $\theta = (\theta_1, \theta_2, \theta_3, ...)$ est la probabilité (ou la densité de probabilité) d'observer les données $(X_i)_{i=1,2,...,n}$ vue comme une fonction de θ :

- X discrète :

$$L(\theta) = \mathbb{P}\left[X_1 = x_1 \cap X_2 = x_2 \cap ... \cap X_n = x_n; \theta\right]$$

- X continue :

$$L(\theta) = h(x_1, x_2, ... x_n; \theta)$$

Vraisemblance - Cas de données indépendantes

- X discrète :

$$L(\theta) = \mathbb{P}[X_1 = x_1; \theta] \times \mathbb{P}[X_2 = x_2; \theta] \times \ldots \times \mathbb{P}[X_n = x_n; \theta]$$

et

$$\ln L(\theta) = \ln \mathbb{P}[X_1 = x_1; \theta] + \ln \mathbb{P}[X_2 = x_2; \theta] + \ldots + \ln \mathbb{P}[X_n = x_n; \theta]$$

- X continue :

$$L(\theta) = f_1(x_1; \theta) \times f_2(x_2; \theta) \times \ldots \times f_n(x_n; \theta)$$

et

$$\ln L(\theta) = \ln f_1(x_1; \theta) + \ln f_2(x_2; \theta) + \ldots + \ln f_n(x_n; \theta)$$

Inférence bayésienne

- Paramètres : θ
- Données : D

Théorème de Bayes :

$$\mathbb{P}[\theta \mid D] = \mathbb{P}[\theta] \; \frac{\mathbb{P}[D \mid \theta]}{\mathbb{P}[D]}$$

- Prior : $\mathbb{P}[\theta]$
- Posterior : $\mathbb{P}[\theta \mid D]$
- Likelihood : $\mathbb{P}[D \mid \theta]$
- Vraisemblance marginale : $\mathbb{P}[D]$

Symboles courants dans la littérature

- Paramètres : θ
- Données : x

Théorème de Bayes :

$$\pi(\theta \mid x) \propto \pi(\theta)\, f(x \mid \theta)$$

- Prior : $\pi(\theta)$
- Posterior : $\pi(\theta \mid x)$
- Likelihood : $f(x \mid \theta)$
- Vraisemblance marginale : ... constante ne dépendant pas de θ

Bayes et MCMC

- Approche bayésienne : caractériser la loi a posteriori $\pi(\theta \mid x)$
- La vraisemblance marginale est souvent compliquée à calculer
- MCMC : échantillonner dans une loi cible $\phi(z)$ en utilisant une loi de proposition $g(z'/z)$:

$$p = \frac{\phi(z')}{\phi(z)} \frac{g(z \mid z')}{g(z' \mid z)}$$

- Bayes et MCMC :

$$\phi(z) = \pi(\theta \mid x) \propto \pi(\theta)\, f(x \mid \theta)$$

Donc :

$$p = \frac{\pi(\theta')\, f(x \mid \theta')}{\pi(\theta)\, f(x \mid \theta)} \frac{g(\theta \mid \theta')}{g(\theta' \mid \theta)}$$

Bayes quand on sait pas calculer la vraisemblance $f(x \mid \theta)$

- Approche bayésienne : caractériser la loi a posteriori $\pi(\theta \mid x)$
- Simulation de couples (θ_S, x_S) :
  - Simuler θ_S dans le prior $\pi(\theta)$
  - Simuler x_S dans la loi $f(x \mid \theta)$
- Si on ne retient que les couples (θ_S, x_S) où $x_S = x$, on obtient un échantillon tiré dans $\pi(\theta \mid x)$