L3 Linguistique Informatique
Université Paris 7

Marie Candito
Jeudi 12 nov 2009

Probabilités et statistiques pour le TAL

Contrôle des connaissances - Correction

Recommandations générales : utilisez au maximum les notations vues en cours. Pour les résultats, inutile de développer/simplifier les factoriels. Pour les probabilités, définissez soigneusement l'ens. fondamental sur lequel travailler.

1 Cours

1.1. Donner les avantages et les inconvénients pour un linguiste de travailler par introspection versus travailler sur corpus.

	Avantages	**Inconvénients**
introspection	- accès à des données inacceptables - accès immédiat aux constructions que l'on étudie	- données pas nécessairement attestées - difficile de quantifier les degrés d'acceptabilité - ne permet pas d'intégrer des informations quantitatives
corpus	- données attestées, éventuellement auxquelles on n'aurait pas pensé par introspection - permettent de quantifier les degrés d'acceptabilité de constructions - fournissent des informations quantitatives, qui sont maintenant considérées comme faisant partie de la compétence (au sens large) d'un locuteur : par ex. un mot rare a un « effet » différent d'un mot fréquent	- parcellaires - pas d'accès aux données agrammaticales, et donc argument de la « pauvreté du stimulus » : corpus sensés être insuffisants à eux seuls pour capturer la compétence - la probabilité de rencontrer un énoncé ne peut pas refléter l'acceptabilité : les énoncés acceptables rares vont être très difficiles à étudier en corpus

1.2. Définissez ce qu'est un arrangement de r objets parmi n. Donnez le nombre d'arrangements de r objets parmi n, en justifiant ce résultat.

=> On a n possibilités pour le premier élément, (n-1) pour le second, (n-2) pour le troisième etc... On obtient par le principe fondamental de dénombrement : n(n-1)(n-2)...(n-r+1), qui peut s'écrire n!/(n-r)!

1.3. Définissez ce qu'est une combinaison de r objets parmi n. Donnez le nombre de combinaisons de r objets parmi n. Justifiez ce résultat en partant du nombre d'arrangements.

=> Une combinaison est un ensemble de r objets pris parmi n. C'est l'équivalent non ordonné des arrangements. Si on considère les n!/(n-r)! arrangements de r objets parmi n, on peut faire des « paquets » d'arrangements, qui sont identiques modulo l'ordre des éléments. Chaque paquet correspond donc exactement à une combinaison. En outre, chacun de ces paquets contient r! éléments, i.e. les r! façons d'arranger les r éléments. On obtient ainsi le nb de combinaisons en divisant n!/(n-r)! par r!

On peut écrire tout evt comme l'union d'un certain nb d'evts élémentaires.

1.4. Donnez la définition axiomatique des probabilités.

=> voir cours

1.5. Dans le cas d'un ensemble fondamental S à évènements élémentaires équiprobables, si E est un événement quelconque sur S, combien vaut P(E) ?

=> En notant N(E) le nombre d'éléments de l'événement E, on aura P(E) = N(E) / N(S)

Evènement → tout sous-ensemble de l'ensemble fondamental.
Evt élémentaire → 1 seule issue

2 Exercices

2.1. Un enfant a appris 10 mots d'italien : 2 déterminants, 4 noms et 4 verbes.

a) Combien peut-il former de phrases de la forme Det N V Det N (dans le cas où chaque mot peut être utilisé plus d'une fois) ?

=> On peut décomposer l'expérience « former une phrase de la forme Det N V Det N » en une suite de plusieurs : « choisir un Det parmi 2», « choisir un N parmi 4» etc... Comme on peut réutiliser chaque mot, on a toujours 2 det, 4 noms et 4 verbes possibles. D'où d'après le principe fondamental de dénombrement :

nombre de telles phrases = 2 x 4 x 4 x 2 x 4 = 256

b) Combien de séquences de mots sont formables (en utilisant chaque mot une fois, et sans considérer les contraintes syntaxiques sur les séquences) ?

=> Il s'agit du nombre de permutations des 10 mots = 10!

c) Combien de séquences de catégories sont formables (en utilisant chaque mot une fois, et sans considérer les contraintes syntaxiques sur les séquences) ?

=> Il s'agit du nombre de permutations des 10 mots, dont respectivement 2, 4, et 4 forment des paquets indiscernables. D'où la réponse : 10! / 2!4!4!

2.2. Un professeur organise un concours de poésie dans sa classe de 12 élèves. Combien de façons a-t-il de choisir 3 équipes de 4 élèves ?

=> Commençons par considérer le cas où l'ordre des équipes est pertinent : on a alors C_{12}^4 façons de choisir les membres de la 1ere équipe, puis C_8^4 façons de choisir la seconde, ce qui dans le même temps choisit la troisième. Donc on aura $C_{12}^4 C_8^4$ façons d'obtenir des triplets ordonnés d'équipes.

Mais dans la question posée, peu importe l'ordre des équipes, dans le résultat précédent on a des triplets d'équipes, qui forment des paquets de 3! triplets équivalents pour peu que l'ordre ne soit plus considéré. D'où le résultat : $C_{12}^4 C_8^4 / 3!$

2.3. Le même enfant n'a toujours que 10 mots d'italien. Il énumère au hasard les 10 mots qu'il connaît.

a) Quelle est la probabilité d'une séquence donnée ?

=> L'ens. fondamental est le nb de permutations des 10 mots, il y en a 10!
L'énumération se faisant au hasard, chaque évènement élémentaire (i.e. ici chaque permution) est équiprobable. La proba d'une des permutations est donc 1 / 10!

b) Quelle est la probabilité que le deuxième mot soit un verbe ?

=> On a 4 façons de choisir le verbe apparaissant en deuxième mot, et 9! façons d'ordonner les autres mots. D'où le résultat = 4x9! / 10! = 4/10

2.4. On tire 5 cartes au hasard parmi un jeu de 52. Quelle est la probabilité de l'événement E= « avoir au moins une carte rouge et au moins une carte noire » ? **Indications :** utilisez les évènements N = « il n'y a que des noires » et R = « il n'y a que des rouges ».

a) donnez P(N) et P(R)

=> L'ens. fondamental est composé des C_{52}^{5}combinaisons de 5 cartes parmi 52. L'événement N comporte C_{26}^{5} éléments, i.e. le nombre de façons de choisir 5 cartes parmi les 26 noires. Idem pour R.

D'où P(N) = P(R) = $C_{26}^{5} / C_{52}^{5} = \frac{26!}{5!21!} \frac{5!47!}{52!} = \frac{23 \times 11}{2 \times 2 \times 51 \times 49} = 0,0253...$

b) Quel axiome des probabilités utiliser pour calculer P(N ∪ R) ?

=> Les deux évènements sont mutuellement exclusifs, donc par l'axiome d'additivité :

P(N ∪ R) = P(N) + P(R)

c) Exprimer E en fonction de N et R (et éventuellement complémentaires), utilisez une loi de Morgan, en déduire le résultat...

=> On a E = « avoir au moins une carte rouge » ∩ « avoir au moins une carte noire »

Or « avoir au moins une carte rouge » est exactement « ne pas avoir que des noires », c'est à dire $\overline{N}$

donc on a $E = \overline{N} \cap \overline{R}$, **donc par une des lois de Morgan** $E = \overline{N} \cap \overline{R} = \overline{N \cup R}$,

d'où $P(E) = P(\overline{N \cup R}) = 1 - P(N \cup R)$

et d'après b) on obtient $P(E) = 1 - P(N \cup R) = 1 - P(N) - P(R) \simeq 0,95$,