Ann. sc. math. Québec, 1987, vol. 11, no 2, pp. 343-349

L'ENTROPIE ET LE MEILLEUR ESTIMATEUR CONVEXE D'UNE VARIABLE ALÉATOIRE

Raymond Leblanc et Silviu Guiasu

Summary

The best linear entropic estimator for a normal random variable is determined. For establishing the best convex linear approximation of a finite or denumerable set of random variables we use the principle of maximum entropy.

Résumé

Nous déterminons le meilleur estimateur linéaire entropique pour une variable normale. Pour obtenir la meilleure approximation convexe pour un ensemble fini ou dénombrable de variables aléatoires, nous employons un principe d'entropie maximale.

1. Le meilleur estimateur entropique pour la moyenne d'une variable aléatoire gaussienne

Soient X une variable aléatoire gaussienne $N(\mu_X,\sigma_X^2)$ et $X_1,\ldots,X_n$ un échantillon. Nous dirons que

$$Y = \sum_{i=1}^{n} \alpha_i X_i$$

est le meilleur estimateur linéaire entropique pour μ_X si Y est non biaisé, i.e.

Mots clés: Principe d'information maximale, estimateur convexe, entropie.

$$E(Y) = \mu_X$$

et si $H(Y) = \min\{H(Z);\ Z = \sum_{i=1}^{n} \beta_i X_i\}$ où H est l'entropie de Shannon [1]. Pour une variable aléatoire continue Z, dont la densité est $f(z)$, nous rappelons que l'entropie est définie par $H(Z) = -\int f(z)\, \ell n\, f(z)\, dz$.

Le meilleur estimateur linéaire entropique pour la moyenne est donc l'estimateur linéaire non biaisé qui minimise l'entropie par rapport à l'ensemble des combinaisons linéaires des composantes de l'échantillon.

THÉORÈME 1. *Pour* X *une variable aléatoire gaussienne* $N(\mu_X,\sigma^2)$, *le meilleur estimateur de* μ_X *est*

$$Y = \frac{X_1 + \ldots + X_n}{n}. \tag{1}$$

PREUVE. Considérons un estimateur de la forme

$$Y = \sum_{i=1}^{n} \alpha_i X_i.$$

L'estimateur Y sera non biaisé si et seulement si

$$\sum_{i=1}^{n} \alpha_i = 1. \tag{*}$$

Comme Y est une combinaison linéaire non biaisée, nous aurons

$$Y \sim N(\mu_X,\sigma_Y^2)$$

où $\sigma_Y^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma_X^2$.

L'entropie de cette variable aléatoire est (voir [1], [2])

$$H(Y) = \tfrac{1}{2} + \tfrac{1}{2} + \ell n(2\pi\sigma_Y^2).$$

Dans ce cas, minimiser $H(Y)$ équivaut à minimiser σ_Y^2. Il nous faut donc déterminer

$$\min \sum_{i=1}^{n} \alpha_i^2$$

sous la contrainte (*). La solution est évidemment donnée par

$$\alpha_i = \frac{1}{n}, \qquad (i = 1,\ldots,n).$$

Par conséquent, la meilleure combinaison linéaire entropique correspond bien à (1). □

Ainsi, pour une variable aléatoire normale, minimiser l'entropie est équivalent à minimiser la variance.

2. La meilleure combinaison convexe de variables aléatoires

Soit $X_1,\ldots,X_n$, n variables aléatoires définies sur un même espace de probabilité et ayant des moments du premier et du deuxième ordre (i.e. moyennes et variances finies). Considérons une combinaison convexe

$$Y = \sum_{i=1}^{n} c_i X_i \tag{2}$$

où $c_i \geq 0$, $(i = 1,\ldots,n)$ et

$$\sum c_i = 1. \tag{3}$$

Les coefficients c_i, $(i = 1,\ldots,n)$ expriment la "contribution" de la variable X_i à la combinaison Y. Remarquons que $c = (c_1,\ldots,c_n)$ satisfaisant (3) représente une distribution de probabilité discrète. Pour déterminer la meilleure combinaison convexe (2), celle présentant la contribution la mieux équilibrée des variables $X_1,\ldots,X_n$. Nous introduirons le critère suivant.

Déterminer le vecteur de coefficients $c = (c_1,\ldots,c_n)$ qui est "le plus large possible" tout en respectant des contraintes données a priori. Plus précisément, déterminer le vecteur des coefficients qui maximise l'entropie H

$$H = H(c) = -\sum_{i=1}^{n} c_i \ln c_i \tag{4}$$

et qui est compatible avec des contraintes données. Appliquant ce principe d'entropie maximale, nous construisons une combinaison convexe qui est la moins discriminante parmi toutes celles qui respectent les contraintes.

THÉORÈME 2.

a) *Etant donné aucune contrainte, la meilleure combinaison convexe des variables aléatoires* $X_1,\dots,X_n$ *est la moyenne arithmétique*

$$Y = \frac{X_1 + \dots + X_n}{n}. \tag{5}$$

b) *Si la moyenne de* Y *est donnée,* $E(Y) = \mu$, *la meilleure combinaison convexe de* $X_1,\dots,X_n$ *est*

$$Y = \frac{1}{\Phi(\beta)} \sum_{i=1}^{n} e^{-\beta\mu_i} X_i \tag{6}$$

où μ_i *est la moyenne de la variable* X_i,

$$\Phi(\beta) = \sum_{i=1}^{n} e^{-\mu_i\beta} \tag{7}$$

et β *la solution (unique) de l'équation*

$$\frac{d \ln \Phi}{d\beta} = -\mu. \tag{8}$$

PREUVE. a) S'il n'y a pas de contrainte autre que (3), l'entropie (4) des coefficients est maximale pour la distribution uniforme (voir [1], pp. 3-4)

$$c_i = \frac{1}{n} \qquad (i = 1,\dots,n).$$

b) Si $E(Y) = \mu$ est donné, de (2) nous obtenons

$$\mu = E(Y) = \sum_{i=1}^{n} c_i\, E(X_i) = \sum_{i=1}^{n} c_i\mu_i \tag{9}$$

et la valeur maximale de l'entropie (4) compatible avec les contraintes (3) et (9) est donnée par la distribution de Gibbs (voir [1], p. 296)

$$c_i = \frac{1}{\Phi(\beta)} e^{-\beta\mu_i}, \qquad i = 1,\dots,n \tag{10}$$

où $\Phi(\beta)$ est donné par (7) après avoir substitué (10) dans la contrainte (3). Le nombre β est la solution de l'équation (8) obtenu en introduisant (10) dans (9).

REMARQUES. 1° Si $X_1,\dots,X_n$ sont des variables aléatoires observables et si Z est une variable aléatoire donnée ayant une moyenne $E(Z) = \mu$, la meilleure approximation linéaire de Z par une fonction de ces variables aléatoires observables, compatible avec la moyenne μ est donnée par la distribution de Gibbs (6).

2^o Très souvent la solution de l'équation (8) est un nombre transcendant. Cependant l'équation (8) est de la forme

$$G(\beta) = 0$$

où G est une fonction strictement croissante prenant des valeurs positives et négatives et par conséquent il est facile d'approcher itérativement la solution β. Par exemple, si

$$\mu_1 = 6, \quad \mu_2 = 8, \quad \mu_3 = 12, \quad \mu = 10{,}92$$

nous avons

$$\beta \simeq -0{,}3914263 \ , \quad \Phi(\beta) = 143{,}00675$$

et la distribution de Gibbs dans ce cas est

$$Y = 0{,}0732164\ X_1 + 0{,}1601757\ X_2 + 0{,}7666079\ X_3.$$

THÉORÈME 3. *Pour* $X_1, X_2, \ldots,$ *une suite de variables aléatoires telles que*

$$\mu_k = E(X_k) = ka, \qquad k = 0,1,\ldots \quad (a > 0),$$

Y *la meilleure combinaison convexe de ces variables, compatible avec la moyenne* $E(Y) = \mu$ *est donnée par*

$$Y = \sum_{k=0}^{\infty} \frac{a\,\mu^k}{(a+\mu)^{k+1}}\, X_k. \tag{11}$$

PREUVE. Nous devons maximiser

$$H = -\sum_{k=0}^{\infty} c_k \,\ell n\, c_k$$

sous la contrainte

$$1 = \sum_{k=0}^{\infty} c_k, \qquad c_k \geq 0 \quad (k = 0,1,\ldots), \tag{12}$$

(13) $$\mu = \sum_{k=0}^{\infty} c_k \mu_k = \sum_{k=0}^{\infty} c_k ka.$$

En utilisant les procédures habituelles pour résoudre ce problème variationnel, nous obtenons

(14) $$c_k = e^{-\alpha-\beta\mu k} \qquad k = 0,1,\ldots$$

où α et β sont les deux multiplicateurs de Lagrange. En introduisant (14) dans (12) et (13), nous obtenons

(15) $$c_k = \frac{1}{\Phi(\beta)} e^{-\beta\mu_k}, \qquad k = 0,1,\ldots$$

$$\Phi(\beta) = \sum_{k=0}^{\infty} e^{-\beta ka} = \frac{1}{1 - e^{-\beta a}}$$

$$\mu = -\frac{\Phi'(\beta)}{\Phi(\beta)} = \frac{ae^{-a\beta}}{1 - e^{-a\beta}}$$

i.e.

(16) $$\Phi(\beta) = \frac{a+\mu}{a}, \qquad e^{-a\beta} = \frac{\mu}{a+\mu}.$$

De (15) et (16), on tire

$$c_k = \frac{a\mu^k}{(a+\mu)^{k+1}} \qquad (k = 0,1,\ldots)$$

et en substituant ces valeurs dans

$$Y = \sum c_k X_k,$$

nous obtenons (11).

Références

[1] GUIASU, S., *Information Theory with Applications*, McGraw-Hill, New York-Toronto-London-Düsseldorf, 1977.

[2] RADHAKRISHNA, RAO C., *Linear Statistical Inference and its Applications*, Second Edition, Wiley, New York-London-Sydney-Toronto, 1965.

Raymond Leblanc
Département de mathématiques
et d'informatique
Université du Québec à Trois-Rivières
C.P. 500, Trois-Rivières
Québec G9A 5H7

Silviu Guiasu
Department of Mathematics
York University
4700 Keele Street
Downsview, Ontario
M3J 1P3

Manuscrit reçu le 14 mars 1986.
Revision le 10 septembre 1986.