Μέση τιμή, διασπορά και συνδιακύμανση
Περίληψη

Αντώνης Οικονόμου
aeconom@math.uoa.gr

31 Μαρτίου 2010

Αν οι X και Y είναι διακριτές τυχαίες μεταβλητές με από κοινού συνάρτηση πιθανότητας $p(x, y)$ και $g(x, y)$ είναι μια συνάρτηση, τότε η μέση τιμή της τυχαίας μεταβλητής $g(X, Y)$ μπορεί να υπολογιστεί από τη σχέση

$$E[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y).$$

Ανάλογα, αν οι X και Y είναι συνεχείς τυχαίες μεταβλητές με από κοινού συνάρτηση πυκνότητας πιθανότητας $f(x, y)$ και $g(x, y)$ είναι μια συνάρτηση, τότε η μέση τιμή της τυχαίας μεταβλητής $g(X, Y)$ μπορεί να υπολογιστεί από τη σχέση

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy.$$

Χρησιμοποιώντας αυτές τις σχέσεις μπορούμε να δούμε ότι ισχύει η σχέση

$$E[X+Y] = E[X] + E[Y],$$

που γενικεύεται στη σχέση

$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i].$$

Προσοχή όμως! Ενώ ισχύει ότι η μέση τιμή του αθροίσματος τυχαίων μεταβλητών ισούται με το άθροισμα των μέσων τιμών, γενικά δεν ισχύει ότι η μέση τιμή του γινομένου τυχαίων μεταβλητών ισούται με το γινόμενο των μέσων τιμών. Όμως:

$$X, Y \quad \text{ανεξάρτητες} \quad \Rightarrow E[XY] = E[X]E[Y]$$

και γενικότερα

$$X_1, X_2, \ldots, X_n \quad \text{ανεξάρτητες} \quad \Rightarrow E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E[X_i].$$

Αν X και Y είναι ανεξάρτητες τυχαίες μεταβλητές τότε και οποιεσδήποτε συναρτήσεις τους $g(X)$ και $h(Y)$ είναι ανεξάρτητες και επομένως θα έχουμε και τη συνεπαγωγή

$$X, Y \quad \text{ανεξάρτητες} \quad \Rightarrow E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Βέβαια χρειάζεται και πάλι προσοχή. Οι αντίστροφες των παραπάνω τριών συνεπαγωγών δεν ισχύουν.

Η **συνδιακύμανση** δυο τυχαίων μεταβλητών X και Y ορίζεται να είναι η ποσότητα

$$Cov[X,Y] = E[(X - E[X])(Y - E[Y])].$$

Δυο τυχαίες μεταβλητές X και Y με $Cov[X,Y]=0$ λέγονται **ασυσχέτιστες**. Είναι εύκολο να δούμε ότι ανεξάρτητες μεταβλητές είναι πάντα ασυσχέτιστες αλλά το αντίστροφο δεν ισχύει. Για τη συνδιακύμανση ισχύουν οι παρακάτω σχέσεις:

$$\begin{aligned}
Cov[X,Y] &= Cov[Y,X], \\
Cov[X,X] &= Var[X], \\
Cov[aX+b,Y] &= aCov[X,Y], \\
Cov[\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j] &= \sum_{i=1}^{n}\sum_{j=1}^{m} Cov[X_i,Y_j].
\end{aligned}$$

Η διασπορά του αθροίσματος τυχαίων μεταβλητών δεν ισούται γενικά με το άθροισμα των διασπορών τους. Για τον υπολογισμό χρειάζεται να γνωρίζουμε και τις συνδιακυμάνσεις μεταξύ των μεταβλητών. Πιο συγκεκριμένα ισχύει η σχέση

$$Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + 2\sum\sum_{i<j} Cov[X_i,X_j].$$

Από τη σχέση αυτή παίρνουμε τη συνεπαγωγή

$$X_1, X_2, \ldots, X_n \quad \text{ανεξάρτητες} \quad \Rightarrow Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i].$$

Ο **συντελεστής (γραμμικής) συσχέτισης** δυο τυχαίων μεταβλητών X και Y είναι ένα ε