Εισαγωγή στην Οικονομετρία

Άγγελος Αλεξόπουλος
angelos@aueb.gr
ΚΩΔ: *PROPTYXIAKA*1404

Διάλεξη 6-7

Εαρινό Εξάμηνο 2024

Διμεταβλητές Τυχαίες Μεταβλητές

- Ένα διμεταβλητό τυχαίο διάνυσμα είναι μια συνάρτηση από το χώρο δειγματοληψίας S στον $\mathbb{R}^2$, δηλαδή στον δισδιάστατο Ευκλείδειο χώρο.
- Οι συντεταγμένες ενός τυχαίου διανύσματος δεν είναι απαραίτητα όλες του ίδιου τύπου (δηλαδή μπορεί να είναι κατηγορικές, διακριτές, συνεχείς, ή οποιασδήποτε συνδυασμός).

Διακριτά Δισδιάστατα Τυχαία Διανύσματα

- Το τυχαίο διάνυσμα (X, Y) ονομάζεται **διακριτό** τυχαίο διάνυσμα αν παίρνει τιμές σε ένα μετρήσιμο (πεπερασμένο ή άπειρο) σύνολο. Τότε η συνάρτηση $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$ που ορίζεται από:

$$f_{X,Y}(x, y) = f(x, y) = P(X = x, Y = y)$$

ονομάζεται από κοινού πιθανότητα **μάζας** ή από κοινού συνάρτηση μάζας πιθανοτήτων του (X, Y), αν:

$$f(x, y) \geq 0 \quad \text{και} \quad \sum_{(x,y)\in\mathbb{R}^2} f(x, y) = P((X, Y) \in \mathbb{R}^2) = 1$$

- Για A, υποσύνολο του $\mathbb{R}^2$, έχουμε:

$$P((X, Y) \in A) = \sum_{(x,y)\in A} f(x, y)$$

Διακριτά Δισδιάστατα Τυχαία Διανύσματα

- **Κοινή Αναμενόμενη Τιμή**:
  Αν $g : \mathbb{R}^2 \to \mathbb{R}$ τότε
  $$E[g(X,Y)] = \sum_{(x,y)\in\mathbb{R}^2} g(x,y)f(x,y)$$
  εφόσον υπάρχει.

- **Περιθώριες πιθανότητες μάζας**:
  Έστω (X, Y) ένα διακριτό δισδιάστατο τυχαίο διάνυσμα με κοινή συνάρτηση μάζας πιθανοτήτων $f(x, y)$. Τότε οι μονοδιάστατες συναρτήσεις πιθανοτήτων του X και του Y δίνονται από:
  $$\begin{aligned} f_X(x) &= P(X=x) = \sum_{y\in\mathbb{R}} f(x,y) \\ f_Y(y) &= P(Y=y) = \sum_{x\in\mathbb{R}} f(x,y) \end{aligned}$$

Συνεχή Δισδιάστατα Τυχαία Διανύσματα

- Μια συνάρτηση $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$ ονομάζεται από κοινού συνάρτηση **πυκνότητας** πιθανότητας ή **από κοινού συνάρτηση πυκνότητας πιθανότητας (σ.π.π.)** του **συνεχούς** δισδιάστατου τυχαίου διανύσματος (X, Y) αν για κάθε $A \subset \mathbb{R}^2$,

$$P((X,Y) \in A) = \iint_A f(x,y)dxdy$$

- Η $f(x,y)$ θα είναι μια δισδιάστατη συνάρτηση πυκνότητας αν

$$f(x,y) \geq 0 \quad \text{και} \quad \iint_{\mathbb{R}^2} f(x,y) = 1$$

και το αντιστροφο ισχύει.

Συνεχή Δισδιάστατα Τυχαία Διανύσματα

- **Από κοινού αναμενόμενη τιμή**: Αν $g : \mathbb{R}^2 \to \mathbb{R}$ τότε

$$E[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y) f(x,y)$$

  εφόσον υπάρχει.

- **Περιθώρια σ.π.π.**: Έστω (X, Y) ένα συνεχές δισδιάστατο τυχαίο διάνυσμα με κοινή συνάρτηση πυκνότητας πιθανότητας $f(x,y)$. Τότε οι μαργιναλ πδφς του X και του Y δίνονται από:

$$\begin{aligned} f_X(x) &= \int_{y \in \mathbb{R}} f(x,y) \\ f_Y(y) &= \int_{x \in \mathbb{R}} f(x,y) \end{aligned}$$

Δεσμευμένες κατανομές

Δεσμευμένη πιθανότητα μάζας.: Έστω (X, Y) ένα **διακριτό** διμεταβλητό τυχαίο διάνυσμα με από κοινού συνάρτηση πιθανότητας μάζας $f(x, y)$ και μονοδιάστατες συναρτήσεις πιθανοτήτων: $f_X(x)$ και $f_Y(y)$.

- Για οποιοδήποτε x τέτοιο ώστε $P(X = x) = f_X(x) > 0$, η δεσμευμενη πιθανότητα μάζας του Y δεδομένου ότι $X = x$ είναι η συνάρτηση του y:

$$f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}$$

- Για οποιοδήποτε y τέτοιο ώστε $P(Y = y) = f_Y(y) > 0$, η δεσμευμενη πιθανότητα μάζας του X δεδομένου ότι $Y = y$ είναι η συνάρτηση του x:

$$f(x|y) = P(X = x|Y = y) = \frac{f(x, y)}{f_Y(y)}$$

Δεσμευμένες κατανομές

Δεσμευμένες σ.π.π.: Έστω (X, Y) ένα **συνεχές** διμεταβλητό τυχαίο διάνυσμα με κοινή συνάρτηση σ.π.π. $f(x, y)$ και περιθώριες σ.π.π.: $f_X(x)$ και $f_Y(y)$.

- Για οποιοδήποτε x τέτοιο ώστε $f_X(x) > 0$, η δεσμευμένη σ.π.π. του Y δεδομένου ότι $X = x$ είναι η συνάρτηση του y:

$$f(y|x) = \frac{f(x, y)}{f_X(x)}$$

- Για οποιοδήποτε y τέτοιο ώστε $f_Y(y) > 0$, η δεσμευμένη σ.π.π. του X δεδομένου ότι $Y = y$ είναι η συνάρτηση του x:

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Δεσμευμένες κατανομές - Ιεραρχική Μοντελοποίηση

- **Δεσμευμένη Αναμονόμενη τιμή**:

$$E[X|Y=y] = \int_{\mathbb{R}} x f(x|y) dx$$

  εφόσον υπάρχει.

- **Δεσμευμένη Διακύμανση**:

$$Var[X|Y=y] = \int_{\mathbb{R}} (x - E[X|Y=y])^2 f(x|y) dx$$

  εφόσον υπάρχει.

- Παρατηρήστε ότι τόσο το $E[X|Y=y]$ όσο και το $Var[X|Y=y]$ είναι συναρτήσεις του y.

Ιεραρχική Μοντελοποίηση

Θεώρημα:

$$E[X] = E[E[X|Y]] \Rightarrow E_X[X] = E_Y[E_{X|Y}[X|Y]]$$

$$Var[X] = E[Var[X|Y]] + Var[E[X|Y]]$$

Απόδειξη:

$$\begin{aligned} E[E[X|Y]] &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} x\, f(x|y) dx \right] f_2(y) dy = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} x\, f(x|y)\, f_2(y) dy \right] dx \\ &= \int_{\mathbb{R}} \left[x \int_{\mathbb{R}} f(x,y) dy \right] dx = \int_{\mathbb{R}} x\, f_1(x) dx = E[X] \end{aligned}$$

$$
\begin{aligned}
Var[X|Y=y] &= \int_{\mathbb{R}} (x - E[X|Y=y])^2 \, f(x|y)dx \\
\Rightarrow E\left[Var[X|Y=y]\right] &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} (x - E[X|Y=y])^2 \, f(x|y)dx \right] f_2(y)dy \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} [x - E(X|Y)]^2 \, f(x,y)dxdy \\
&= E\left[[X - E(X|Y)]^2 \right]
\end{aligned}
$$

$$
E\left[Var[X|Y=y]\right] = E\left| [X - E(X|Y)]^2 \right| \quad \textbf{(I)}
$$

$$
\begin{aligned}
Var(X) &= \int_{\mathbb{R}} \int_{\mathbb{R}} (x - E[X])^2 f(x,y)dxdy = E\left[(X - E[X])^2 \right] \\
&= E\left[(X - E[X|Y] + E[X|Y] - E[X])^2 \right] \\
&= E\left[(X - E[X|Y])^2 \right] + E\left[(E[X|Y] - E[X])^2 \right] \\
&\quad +2E\left\{ (X - E[X|Y])(E[X|Y] - E[X]) \right\} = \textbf{(II)}
\end{aligned}
$$

$$
\begin{aligned}
E\left[[X - E(X|Y)]^2\right] &= E\left[Var[X|Y = y]\right] \quad \text{by } \mathbf{(I)} \\
E\left[(E[X|Y] - E[X])^2\right] &= E\left[(E[X|Y] - E[E[X|Y]])^2\right] = Var[E[X|Y]]
\end{aligned}
$$

Πρέπει μόνο να δείξουμε ότι

$$\begin{aligned} & E\{(X - E[X|Y])(E[X|Y] - E[X])\} \\ = \; & E[E\{(X - E[X|Y])(E[X|Y] - E[X])|Y\}] = 0 \end{aligned}$$

Τότε η (II) θα μας δίνει το ζητούμενο. Έστω

$$A = E\{(X - E[X|Y])(E[X|Y] - E[X])|Y\}$$

Τότε

$$\begin{aligned} A &= E\{(X - E[X|Y])(E[X|Y] - E[X])|Y\} \\ &= [E[X|Y] - E[X]]\,[E\{(X - E[X|Y])|Y\}] \\ &= [E[X|Y] - E[X]] \times 0 \\ &= 0 \end{aligned}$$

Ανεξαρτησία

Θεώρημα Οι X και Y είναι ανεξάρτητες αν και μόνον αν ισχύει μία από τις παρακάτω περιπτώσεις:

α. $F(x, y) = F_X(x)F_Y(y) \ \ \forall x, y$

β. $p(x, y) = p_X(x)p_Y(y)$, όταν οι X και Y είναι διακριτές.

γ. $f(x, y) = f_X(x)f_Y(y)$, όταν οι X και Y είναι συνεχείς.

- Εάν οι X και Y είναι ανεξάρτητες, τότε η δεσμευμενη κατανομή του $Y|X$ είναι ίδια με την μαργιναλ κατανομή του Y, δηλαδή $f(y|x) = f_Y(y)$.

- Οι X και Y είναι ανεξάρτητες αν και μόνον αν:

$$E[h_1(X)h_2(Y)] = E[(h_1(X))]E[(h_2(Y))]$$

για όλες τις συναρτήσεις $h_1(.)$ και $h_2(.)$ για τις οποίες τα ολοκληρώματα υπάρχουν.

Covariance/**Συνδιακύμναση**

- Έστω (X, Y) ένα ζευγάρι τυχαίων μεταβλητών με $\mu_X = E[X]$ και $\mu_Y = E[Y]$. Τότε η **ὂαριανςε** του X και Y είναι ο αριθμός που ορίζεται από:

$$Cov(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right] = E[XY] - E[X]E[Y]$$

- X και Y είναι ανεξάρτητες $\Rightarrow Cov(X, Y) = 0$ το αντίστροφο **δεν** είναι αληθές.

- Εάν X και Y είναι οποιεσδήποτε δύο τυχαίες μεταβλητές και a και b είναι οποιεσδήποτε δύο σταθερές, τότε:

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

Correlation/**Συσχέτιση**

- Έστω (X, Y) ένα ζευγάρι τυχαίων μεταβλητών με $0 < Var[X] < +\infty$ και $0 < Var[Y] < +\infty$. Τότε ο συντελεστής **ςορρελατιον/συσχέτισης** ρ του X και του Y είναι ο αριθμός που ορίζεται από:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var[X]}\sqrt{Var[Y]}}$$

- Για $ac \neq 0$ ισχύει:

$$\rho(aX + b, cY + d) = \frac{ac}{|ac|}\rho(X, Y)$$

- Γενικά: $-1 \leq \rho(X, Y) \leq 1$ $\rho(X, Y)^2 = 1 \Leftrightarrow Y = kX$ για κάποια σταθερά k.
- Αν οι X και Y είναι ανεξάρτητες, $\Rightarrow$ X και Y είναι μη συσχετισμένες/υνςορρελατεδ ($\rho(X, Y) = 0$). Η αντίστροφη πρόταση δεν είναι αληθής.

Τυχαία Διανύσματα

- Ένα διάνυσμα $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots\ X_n \end{bmatrix}$ είναι ένα τυχαίο διάνυσμα εάν για οποιοδήποτε $\boldsymbol{t} = \begin{bmatrix} t_1 \\ t_2\ \vdots\ t_n \end{bmatrix} \in \mathbb{R}^n$ έχουμε: $\boldsymbol{t}'X$ να είναι μια πραγματική τυχαία μεταβλητή.
- Εάν $\boldsymbol{t}_i = \begin{bmatrix} 0 \vdots t_i \vdots 0 \end{bmatrix} \in \mathbb{R}^n$ έχουμε: τότε $X_i = \boldsymbol{t}_i'X$ είναι μια συνιστώσα (ή προβολή).

Τυχαία Διανύσματα

- Ένα διάνυσμα $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \ X_n \end{bmatrix}$ είναι ένα τυχαίο διάνυσμα εάν για οποιοδήποτε $\boldsymbol{t} = \begin{bmatrix} t_1 \\ t_2 \ \vdots \ t_n \end{bmatrix} \in \mathbb{R}^n$ έχουμε: $\boldsymbol{t}'X$ να είναι μια πραγματική τυχαία μεταβλητή.
- Εάν $\boldsymbol{t}_i = \begin{bmatrix} 0 \ \vdots \ t_i \ \vdots \ 0 \end{bmatrix} \in \mathbb{R}^n$ έχουμε: τότε $X_i = \boldsymbol{t}_i'X$ είναι μια συνιστώσα (ή προβολή).
- Συνάρτηση πολυμεταβλητής αθροιστικής κατανομής :

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

Τυχαία Διανύσματα

- **Συνεχή τυχαία διανύσματα:** Το X είναι συνεχές εάν:

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(u_1, \ldots, u_n) du_1 \ldots du_n$$

  που:

$$f(u_1, \ldots, u_n) \geq 0 \quad \text{και} \quad \int_{\mathbb{R}^n} f(u_1, \ldots, u_n) du_1 \ldots du_n = 1$$

- **Αναμενόμενη τιμή:** Αν $g : \mathbb{R}^n \to \mathbb{R}$ τότε:

$$E[g(\mathrm{X})] = \int_{\mathbb{R}^n} g(\mathrm{X}) f(\mathrm{X}) dx$$

  εφόσον το ολοκλήρωμα υπάρχει.

Ανεξαρτησία

- Οι συντεταγμένες του τυχαίου διανύσματος X: $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες αν και μόνον αν:

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

- Οι συντεταγμένες του τυχαίου διανύσματος X: $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες αν και μόνον αν:

$$E\left[\prod_{i=1}^{n} h(X_i)\right] = \prod_{i=1}^{n} E[(h_i(X_i))]$$

για όλες τις συναρτήσεις $h_i(.)$, για τις οποίες τα ολοκληρώματα υπάρχουν.

Τυχαίο Δείγμα

- Οι τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$ ονομάζονται ένα **τυχαίο δείγμα** μεγέθους n αν:
  (α) Οι $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες
  (β) Οι περιθώριες της κάθε X_i είναι η ίδια συνάρτηση $f(x)$.

- Αν οι $X_1, X_2, \ldots, X_n$ είναι τυχαίο δείγμα, τότε τις ονομάζουμε *independent identically distributed* (*iid*) τυχαίες μεταβλητές και τότε ισχύει:

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Τυχαίο Δείγμα

- Η συνάρτηση $f(x_1, \ldots, x_n|\theta)$ είναι γνωστή ως **πιθανοφάνεια**, και συμβολίζεται ως $L(\theta|\underline{x}) = L(\theta|x_1, \ldots, x_n)$, και για ένα συγκεκριμένο σύνολο δεδομένων $(x_1, x_2, \ldots, x_n)$ είναι απλά μια συνάρτηση της άγνωστης παραμέτρου θ, δηλαδή $\ell(\theta)$.

- **Σημείωση**: X_i υποδηλώνει την τυχαία μεταβλητή ενώ x_i είναι η τιμή της τυχαίας μεταβλητής σε ένα τυχαίο δείγμα (δηλαδή $X_i = x_i$)

Στατιστικές Συναρτήσεις

- $X_1, \ldots, X_n$ είναι ένα τυχαίο δείγμα μεγέθους n από έναν πληθυσμό και έστω $T(x_1, \ldots, x_n)$ μια πραγματική (ή διανυσματική) συνάρτηση η οποία έχει σαν πεδίο ορισμού τον δειγματικό χώρο των $(X_1, \ldots, X_n)$. Τότε η τυχαία μεταβλητή (ή διάνυσμα) $Y = T(X_1, \ldots, X_n)$ ονομάζεται **Στατιστική Συνάρτηση**.

- Δεδομένου ότι η τιμή του Y θα ποικίλλει από δείγμα σε δείγμα, μπορούμε να βρούμε την κατανομή πιθανότητας, την οποία θα ονομάσουμε **δειγματική κατανομή** του Y.

- Μια στατιστική **δεν** μπορεί να είναι μια συνάρτηση της παραμέτρου.

- Δύο από τις πιο δημοφιλείς στατιστικές συναρτήσεις είναι:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Θεώρημα : Έαν $X_1, \ldots, X_n$ είναι ένα τυχαίο δείγμα μεγέθους n από έναν πληθυσμό και $g(x)$ συνάρητση τ.ω. $E[g(X_1)]$ και $Var[g(X_1)]$ ορίζονται, τότε

$$E\left[\sum_{i=1}^{n} g(X_i)\right] = nE[g(X_1)]$$

και

$$Var\left[\sum_{i=1}^{n} g(X_i)\right] = nVar[g(X_1)]$$

Θεώρημα : Έαν $X_1, \ldots, X_n$ είναι ένα τυχαίο δείγμα μεγέθους n από έναν πληθυσμό με $E[X_i] = \mu$ και $Var[X_i] =^2 < \infty$ τότε

$$E\left[\overline{X}\right] = \mu \qquad V\left[\bar{X}\right] = \frac{\sigma^2}{n} \qquad E\left[s^2\right] = \sigma^2$$

Proof:

$$\begin{aligned} E\left[\overline{X}\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}nE\left[X_1\right] = \mu \\ V\left[\overline{X}\right] &= V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}V\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}nV\left[X_1\right] = \frac{\sigma^2}{n} \end{aligned}$$

$$
\begin{aligned}
E\left[S^2\right] &= E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)\right] = \frac{1}{n-1}\left[nE\left[X_1^2\right] - nE\left[\overline{X}^2\right]\right] \\
&= \frac{1}{n-1}\left[n\left(V\left[X_1\right] + \left(E\left[X_1\right]\right)^2\right) - n\left(V\left[\overline{X}\right] + \left(E\left[\overline{X}\right]\right)^2\right)\right] \\
&= \frac{1}{n-1}\left[n\left(\sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right] \\
&= \frac{1}{n-1}\left[n\sigma^2 + \sigma^2\right] = \frac{(n-1)\sigma^2}{n-1} = \sigma^2
\end{aligned}
$$