Πανεπιστήμιο Πειραιώς
Σχολή Χρηματοοικονομικής και Στατιστικής
Τμήμα Χρηματοοικονομικής και Τραπεζικής Διοικητικής

Πρόγραμμα Μεταπτυχιακών Σπουδών στη
Χρηματοοικονομική και Τραπεζική

Προπαρασκευαστικο Μαθημα Στη Στατιστικη

Μέρος 3ο

3.1 Continuous Probability Distributions

- Continuous Random Variable (r.v.)
  - It can take any value within a given interval:
    - We are not interested in the probability to take a specific value:
      - When Y is continuous r.v., P(Y=y)=0 for every y.
    - We are interested in the probability that Y belongs to an interval (this is, for example, P(Y ∈ (a,b)) where a<b)
    - This probability can be given in terms of a "density function" (συνάρτηση πυκνότητας), *f*(x):

$$P(Y \in (a,b)) = \int_a^b f(x)dx$$

    - This density is the derivative of the Cumulative Distribution Function (αθροιστική συνάρτηση κατανομής) (CDF) of *Y*: $f(x)=F'(x)$. Therefore

$$P(Y \in (a,b)) = \int_a^b f(x)dx = F(b) - F(a).$$

  - Properties of a CDF:
    - $F(-\infty) = 0$
    - $F(\infty) = 1$
    - *F* is non-decreasing
    - *F*(x) = P(*Y*≤x) where x can be any real number, or $-\infty$, or ∞.

3.2 Moments (Ροπες)of Continuous Random Variables

- Expectation (or mean)
  - μ =E[Y]=$\int_{-\infty}^{\infty} yf(y)dy$
- Variance
  - Var(Y)=$\int_{-\infty}^{\infty}(y - E[Y])^2 f(y)dy$
- In general, the k-th moment of Y is given by
  - $\mu_k = E[Y^k] = \int_{-\infty}^{\infty} y^k f(y)dy$
- The k-th central moment of Y is given by
  - $\mu_k^0 = E[(Y - \mu)^k] = \int_{-\infty}^{\infty}(y - \mu)^k f(y)dy$
  - Therefore Var(Y) = μ_2^0

3.3 Normal (or Gaussian) Distribution

- A r.v. Y is normal if its density function is of the form

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

- *f(y)* : height of the distribution at *y.*
- μ : mean of the random variable
- σ : standard deviation of the random variable
- e = 2.718
- π = 3.142
- Notation: $Y \sim N(\mu, \sigma^2)$

3.4 Standard Normal Distribution

- The normal distribution with zero mean and unit variance is called "*standard normal distribution*".
- If Z follows the standard normal distribution, we write *Z~N(0,1)*
- If $Y \sim N(\mu, \sigma^2)$ then the r.v. $Z = \frac{Y-\mu}{\sigma} \sim N(0,1)$
- Z is a transformation of *Y*
- If we know the cumulative distribution of *Z* then we know the cumulative distribution of *Y*
- Example:

  if μ=1 and σ^2=4 and we want to find the probability $P(Y \in (a,b))$ then this equals the probability $P\left(Z \in \left(\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}\right)\right)$

3.4 Standard Normal Distribution (continued)

- Important numbers:
  - $P(-1.96 \leq Z \leq 1.96) = 0.95 = 95\%$
- Note: The density of any normal r.v., *Y*, is symmetric around its mean. Therefore,
  - for every number $\alpha \geq 0$, $P(Y \leq \mu - \alpha) = P(Y \geq \mu + \alpha)$
  - In particular, for $\alpha = 0$, $P(Y \leq \mu) = P(Y \geq \mu) = \frac{1}{2} = 0.5 = 50\%$

3.5 Joint Distribution of Random Variables

- It is usual that in our studies and research we have to deal with more than one random variables at the same time.
  - Example: We would like explore a possible association between smoking of pregnant women and the salary of their children when they reach 30.
  - The random variable that concerns smoking, say X, can take two values (1 if affirmative, 0 otherwise). The random variable that corresponds to the salary, say Y, can be the annual income.
  - For each individual, we have two values (one for each random variable).

3.5 Joint Distribution of Random Variables (continued)

- We say that these variables have a joint (από κοινού) distribution with cumulative distribution function (CDF):
$$F_{X,Y}(x, y) = P(X \leq x \text{ } and \text{ } Y \leq y)$$
- $F_{X,Y}$ has the following properties
  - $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$
  - $F_{X,Y}(\infty, \infty) = 1$
  - $F_{X,Y}$ is increasing in both *x* and *y*.
- Note that $P(X \leq x) = P(X \leq x \text{ } and \text{ } Y \leq \infty)$. We say that the Marginal (περιθώρια) Distribution of *X* is $F_X(x) = F_{X,Y}(x, \infty) = P(X \leq x)$. Similarly, the Marginal Distribution of *Y* is $F_Y(y) = F_{X,Y}(\infty, y) = P(Y \leq y)$.
- The notion of the Joint Distribution is easily extended to more than 2 RVs. So, if $X_1, X_2, X_3, \ldots, X_d$ are RVs, their joint distribution is denoted as
$$F_{X_1,X_2,X_3,\ldots,X_d}(x_1, x_2, x_3, \ldots, x_d)$$
- The marginal distribution of, say, X_2 is then
$$F_{X_2}(x_2) = F_{X_1,X_2,X_3,\ldots,X_d}(\infty, x_2, \infty, \ldots, \infty)$$

3.6 Independence of Random Variables

- If in the previous example were no association between the two random variables, then intuitively we could say that the two random variables are "independent"
- In probabilistic terms, the occurrence of any value of X would not affect any of the probabilities that Y takes a value within any interval (a,b].
- More formally, we can say that X and Y are independent if for any values x and y,

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

3.6 Independence of Random Variables (continued)

- Case 1: Discrete random variables
  - Let the RVs X and Y take the values $\ldots<x_{-2}<x_{-1}<x_0<x_1<x_2<\ldots$, and $\ldots<y_{-2}<y_{-1}<y_0<y_1<y_2<\ldots$, respectively.
  - Their Joint CDF is given by

$$F_{X,Y}(x,y) = \sum_{x_i \le x, y_j \le y} P(X = x_i \text{ and } Y = y_j) = \sum_{x_i \le x, y_j \le y} P(x_i, y_j)$$

$$= \sum_{x_i \le x, y_j \le y} f_{X,Y}(x_i, y_j)$$

where $f_{X,Y}(x_i, y_j)$ is the joint probability (mass) function of X and Y.

  - The marginal probability mass function of X is given by

$$f_X(x_i) = \sum_{y_j \le \infty} f_{X,Y}(x_i, y_j) = \mathrm{P}(\mathrm{X} = x_i)$$

Similarly for Y.

3.6 Independence of Random Variables (continued)

- Case 1: Discrete random variables (continued)
  - The marginal CDF of X is given by

$$F_X(x) = F_{X,Y}(x,\infty) = \sum_{x_i \le x, y_j \le \infty} P\left(X = x_i \ and \ Y = y_j\right)$$

$$= \sum_{x_i \le x, y_j \le \infty} P\left(x_i, y_j\right) = \sum_{x_i \le x, y_j \le \infty} f_{X,Y}\left(x_i, y_j\right)$$

  Similarly we define $F_Y(y)$.

  - We say that the two discrete RVs are **independent** if for every pair (x_i, y_j)

$$f_{X,Y}\left(x_i, y_j\right) = f_X(x_i) f_Y\left(y_j\right)$$

3.6 Independence of Random Variables (continued)

- Case 2: Continuous random variables
  - Let now X and Y be two continuous random variables with joint CDF $F_{X,Y}$. Then there exists a function $f_{X,Y}$ such that
    $$F_{X,Y}(x,y) = P(X \leq x \ and \ Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)dtds$$
  - $f_{X,Y}$ is called "joint probability density function" of X and Y.
  - The marginal probability density function of X is then defined as
    $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,t)dt$$
  - Similarly, the marginal probability density function of Y is defined as
    $$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(s,y)ds$$

3.6 Independence of Random Variables (continued)

- Case 2: Continuous random variables (continued)
  - It can be shown that *X* and *Y* are independent if for every *x* and *y*,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

3.7 χ^2 DISTRIBUTION

- If *Z* is a r.v. and *Z*~*N*(0,1), then Z^2 follows the χ^2 distribution with 1 degree of freedom
- The sum of the squares of *n* independent standard normal random variables follows the Chi-square distribution with *n* degrees of freedom $\chi^2(n)$.
- Properties
  - It is right-skewed (positive skewness) and non-negative
  - Skewness decreases as *n* increases
  - Its mean equals *n*
  - Its mode is *n*-2
  - Its median is approximately *n*-7
  - Its variance is 2*n*

3.8 Student *t*-Distribution

- Let $Z \sim N(0,1)$ and $Y \sim \chi^2(n)$. Then, if Z and Y are independent, we say that the r.v.

$$\tau = \frac{X}{\sqrt{Y/n}}$$

  is the student *t* distribution with *n* degrees of freedom.
- We denote $\tau \sim t(n)$.
- It is leptokurtic due to its "*fatter tails*" than the ones of the normal distribution
- As *n* increases, $t(n)$ approximates the standard normal distribution.

3.9 Fisher's F distribution

- Let X_1 and X_2 two **independent** Chi-square distributions with degrees of freedom n and m, respectively. The random variable

$$F = \frac{X_1/n}{X_2/m} \sim F(n, m)$$

  is called F distribution with degrees of freedom n and m
- Properties
  - F is nonnegative
  - F is right-skewed

Άσκηση 4

Στον παρακάτω πίνακα αναγράφεται η από κοινού συνάρτηση κατανομής πιθανότητας των τυχαίων μεταβλητών *X* και *Y*.

y \ x	0	1	2
0	0.2	0.2	0.2
2	0.1	0.1	0.2

Να εξετάσετε κατά πόσο οι τυχαίες μεταβλητές *X* και *Y* είναι ανεξάρτητες.

Άσκηση 5

Αν δύο τυχαίες μεταβλητές X και *Y* έχουν ως από κοινού συνάρτηση πυκνότητας πιθανότητας την ακόλουθη

$$f_{X,Y} = \begin{cases} 4(x+y^2), & \text{αν } x, y > 0 \text{ και } x + y \leq 1 \\ 0, & \text{διαφορετικά} \end{cases}$$

Να εξεταστεί αν οι *X* και *Y* είναι ανεξάρτητες