Πανεπιστήμιο Πειραιώς
Σχολή Χρηματοοικονομικής και Στατιστικής
Τμήμα Χρηματοοικονομικής και Τραπεζικής Διοικητικής

Πρόγραμμα Μεταπτυχιακών Σπουδών στη
Χρηματοοικονομική και Τραπεζική

Προπαρασκευαστικο Μαθημα Στη Στατιστικη

Μέρος 4ο

4.1 Expectation of a Function of Discrete Random Variables

- Let X and Y be two discrete RVs with joint CDF $F_{X,Y}$ and joint probability mass function $f_{X,Y}$. Let $u(X,Y)$ be a function of these variables. If this function satisfies a quite general condition, then $u(X,Y)$ is also a random variable. Moreover,

$$E[u(X,Y)] = \sum_{x_i \leq \infty, y_j \leq \infty} u(x_i, y_j) f(x_i, y_j)$$

- Simple examples
  - If $u(X,Y) = X$ then

$$E[u(X,Y)] = E[X] = \mu_X = \sum_{x_i \leq \infty, y_j \leq \infty} x_i f(x_i, y_j) = \sum_{x_i \leq \infty} x_i f_X(x_i)$$

  - If $u(X,Y) = (X - \mu_X)(Y - \mu_Y)$ then we have the covariance of X and Y:

$$E[u(X,Y)] = Cov(X,Y) = \sum_{x_i \leq \infty, y_j \leq \infty} (x_i - \mu_X)(y_j - \mu_Y) f(x_i, y_j)$$

  - A special case gives us the variance of X: $Var(X) = Cov(X,X)$

4.2 Expectation of a Function of Continuous Random Variables

- Let X and Y be two continuous RVs with joint CDF $F_{X,Y}$ and joint probability density function $f_{X,Y}$. Let $u(X,Y)$ be a function of these variables. If this function satisfies a quite general condition, then $u(X,Y)$ is also a random variable. Moreover,

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

- Simple examples
  - If $u(X,Y) = X$ then

$$E[u(X,Y)] = E[X] = \mu_X = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dydx = \int_{-\infty}^{\infty} xf_X(x)\,dx$$

  - If $u(X,Y) = (X - \mu_X)(Y - \mu_Y)$ then we have the covariance of X and Y:

$$E[u(X,Y)] = Cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x,y)\,dydx$$

  - A special case gives us the variance of X: $Var(X) = Cov(X,X)$ (similarly for Y)

4.3 Properties of the Expectation

- Let X and Y be two RVs with joint CDF $F_{X,Y}$, then the following hold
- If X and Y have finite expectations, then
  - $E[aX+bY]=aE[X]+bE[Y]$ for any two real numbers a and b.
- If in addition X and Y are independent, then
  - $E[XY]=E[X]E[Y]$

4.4 Covariance of Two Random Variables

- The covariance (συνδιακύμανση ή συνδιασπορά) of two RVs provides an indication of whether these variables increase or decrease together.
- A positive covariance is an indication of such a behavior.
- On the other hand, a negative covariance shows that when the one variable increases we expect that the other decreases.
- A nonzero covariance is an indication of dependence between the random variables because:
- **If X and Y are independent, then**

$$Cov(X,Y)=0.$$

- **On the other hand, we can have two dependent RVs with a zero covariance.**

4.5 Properties of the Covariance of Random Variables

- Let X and Y be two RVs with finite variances. Then
  - $Cov(aX,bY)=abCov(X,Y)$
  - $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$

4.6 Correlation of Random Variables

- As we already explained, the sign of the covariance of two RVs provides information about whether they move together or not.
- On the other hand, the magnitude of the covariance is affected by the magnitude of the variance of each RV.
- In order to identify how strong the relationship of two variables is, we must rescale the covariance, so that it is not affected by the scale of measurements of the two variables.
- For this purpose we use the Correlation:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

where σ_X and σ_Y are the standard deviations of *X* and *Y*, respectively.

4.6 Correlation of Random Variables (continued)

- Properties of the Correlation of two RVs, *X* and *Y*:
  - $-1 \leq Corr(X, Y) \leq 1$
  - $Corr(X, X) = 1 = Corr(Y, Y)$
  - $Corr(X, -X) = -1$
  - If $Corr(X, Y) = 1$ we can be almost sure that the two random variables coincide (*X*=*Y* almost surely)
  - Correlation is invariant to scaling. In other words
    - $Corr(aX, bY) = sgn(ab) Corr(X, Y)$ for any two real numbers *a* and *b*, where $sgn(x)$ denotes the sign of the number *x*.
  - This also means that Correlation is invariant to the units of measurement.

Άσκηση 6 (Συνέχεια Άσκησης 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

(α) Να υπολογίσετε το συντελεστή συσχέτισης $Corr(X,Y)$.
(β) Έστω η τυχαία μεταβλητή $Z=2X+3Y$. Να υπολογίσετε τη $Var(Z)$.

4.7 Conditional Distributions

- In many cases where we start with many random variables, which have a joint CDF, *F*, we reach a point where we know the realizations of some of the variables while we do not know the realizations of the rest of them. If these variables are not independent, the information about the new realizations may be exploited by updating (changing) the probabilities for the events of the “unrealized” random variables.
- In the simplest case, when we have two discrete random variables *X* and *Y*, with joint probability mass function $f_{X,Y}$, we can directly use the definition of conditional probability.
- Specifically, the conditional probability mass function of *Y* given that $X=x_i$ is given by

$$f_Y(y_j|X=x_i) = \frac{f_{X,Y}(x_i, y_j)}{f_X(x_i)}$$

4.7 Conditional Distributions (continued)

- In case that X and Y are both continuous RVs it can be shown that a similar formula holds for the conditional probability density function of Y given that $X = x$:

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

- Note that x (or x_i in the discrete case) must belong in the *support* (στήριγμα) of the RV X, which means that it must hold $f_X(x) \neq 0$ (or $f_X(x_i) \neq 0$ for the discrete case).
- By using the conditional probability density (or mass) function we can obtain the conditional CDF of Y as in the case where we had one RV.

4.7 Conditional Distributions (continued)

- We can use conditional distributions to identify whether two RVs are independent or not. Specifically
  - If *X* and *Y* are discrete, then *X* and *Y* are independent if
    $$f\left(y_j \middle| X = x_i\right) = f_Y\left(y_j\right)$$
    for all x_i in the support of *X.*
  - Similarly, if *X* and *Y* are continuous, then *X* and *Y* are independent if
    $$f(y|X = x) = f_Y(y)$$
    for all *x* in the support of *X.*
- **In order to apply the above criterion, we do not have to calculate the conditional distributions of *Y* for every value of the support of *X*. We obtain the formula of $f\left(y_j \middle| X = x_i\right)$ (or $f(y|X = x)$) maintaining the notation x_i (or *x*). If x_i (or *x*) does not appear in the derived distribution, then the two RVs are independent.**

4.8 CONDITIONAL MOMENTS

- Using the conditional distribution of *Y* we can derive its Conditional Expectation given a value of *X*. Specifically, we have
  $E[Y|X = x_i] = \sum_{y_j \leq \infty} y_j f(y_j|X = x_i)$ for the discrete case, and
  $E[Y|X = x] = \int_{-\infty}^{\infty} y f(y|X = x) dy$ for the continuous case.
- In general, if $g(\cdot)$ is a function which has a "good property" (all the functions you know have it) then
  $E[g(Y)|X = x_i] = \sum_{y_j \leq \infty} g(y_j) f(y_j|X = x_i)$ for the discrete case, and
  $E[g(Y)|X = x] = \int_{-\infty}^{\infty} g(y) f(y|X = x) dy$ for the continuous case.

4.8 Conditional Moments (continued)

- The conditional variance of *Y* is calculated by incorporating the conditional expectation of *Y* instead of its “unconditional” expectation. Specifically $Var(Y|X = x_i) = E[(Y - E[Y|X = x_i])^2|X = x_i]$ for the discrete case, and $Var(Y|X = x) = E[(Y - E[Y|X = x])^2|X = x]$ for the continuous case.

Ασκηση 7 (συνεχεια Ασκησεων 4 και 5)

(Α) 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 υπολογίσετε τη δεσμευμένη συνάρτηση πιθανότητας $f(x_i|Y=y_j)$ και τη δεσμευμένη διακύμανση $Var(X|Y=2)$.

Β) Αν δύο τυχαίες μεταβλητές Χ 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 υπολογιστεί η $Var(Y|X=x)$

4.9 Conditional Expectation as a Random Variable

- In the first part of the previous problem, we had to calculate the conditional expectation of *X* given that *Y*=2. On the other hand, the event *Y*=2 occurs with a specific probability (in fact $P(Y = 2) = 0.4$).
- Therefore, if we do not know the realized value of Y, there is a 40% probability that the conditional expectation of *X* takes the value we have already calculated (i.e. $E[X|Y = 2]$).
- A similar argument concludes that with 60% probability, the conditional expectation of *X* will take the value $E[X|Y = 0]$.
- Therefore, we can expand the notion of the conditional expectation in an "agnostic" way saying that $E[X|Y]$ is a random variable with

$$P(E[X|Y] = E[X|Y = 0]) = 0.6 \text{ and}$$
$$P(E[X|Y] = E[X|Y = 2]) = 0.4$$

4.9 Conditional Expectation as a Random Variable (continued)

- This view of conditional expectations can be applied to cases that involve many RVs, and of course to the case of continuous RVs.
- Main properties of the conditional expectation
  - $E\left[E[Y|X]\right] = E[Y]$ (law of iterated expectations)
  - $E[E[g(X)Y|X]] = E[g(X)Y]$
  - $Var(Y) = E[Var(Y|X)] + var(E[Y|X])$

Άσκηση 8

- Έστω ότι η από κοινού κατανομή των τυχαίων μεταβλητών X και Y είναι τέτοια ώστε
  - $E[X] = 2$
  - $E[X^2] = 16$
  - $E[Y|X] = 2 + X$ και
  - $Var(Y) = 49$

Να βρεθούν οι $Cov(X, Y)$ και $Corr(X, Y)$.

5.1 Εκτιμητριες (Εκτιμητες) Estimators

- Με τον όρο «**Εκτιμήτρια**» ή «**Εκτιμητής**» χαρακτηρίζουμε έναν κανόνα μέσω του οποίου χρησιμοποιόντας τα δεδομένα μας καταλήγουμε στην εκτίμηση μιας (άγνωστης) παραμέτρου ενδιαφέροντος.
- **Κάθε εκτιμήτρια είναι τυχαία μεταβλητή (αφού κάθε τιμή που μπορεί να πάρει εξαρτάται από το δείγμα)**
- **Εκτίμηση** (estimate ή point estimate) είναι η τιμή που προκύπτει από την εφαρμογή του κανόνα αυτού στα δεδομένα που έχουμε.

5.1 Εκτιμητριεσ (Εκτιμητεσ) Estimators (continued)

- **Παράδειγμα:**
  - Έστω ότι το αναμενόμενο ύψος ενός ενήλικα Έλληνα είναι μ.
  - Έστω ότι για ένα οποιοδήποτε τυχαίο δείγμα μεγέθους *n* από το σύνολο του ενήλικου πληθυσμού. Τα αντίστοιχα ύψη τα συμβολίζουμε $X_1, X_2, \ldots, X_n$. Ισχύει ότι $E(X_i) = \mu$ για κάθε *I* (και ότι $\mathrm{Var}(X_i)=\sigma^2$ για κάθε *i* - υποθέτουμε ότι $\sigma^2 < \infty$).
  - Η κλασική εκτιμήτρια (ο κανόνας δηλαδή) από την οποία θα προκύψει η εκτίμηση του μ είναι η $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - **Προσοχή:** τα $X_1, X_2, \ldots, X_n$ είναι τυχαίες μεταβλητές. Όταν πάρουμε ένα συγκεκριμένο τυχαίο δείγμα, τότε συγκεκριμενοποιούνται (πραγματοποιούνται) οι τιμές τους.
    - **Παράδειγμα:** Αν διαλέξουμε 7 ανθρώπους (δηλαδή n=7) με ύψη
      1,57 1,82 1,65 1,78 1,62 1,77 1,68
      τότε
      X_1= 1,57 X_2= 1,82 X_3= 1,65 X_4= 1,78 X_5= 1,62 X_6= 1,77 X_7= 1,68
      Με βάση αυτό το δείγμα, η εκτίμηση του μ είναι

$$\hat{\mu} = \frac{1}{7}\sum_{i=1}^{7} X_i \cong 1{,}70$$

5.2 (Good) Properties of Estimators

- **Unbiasedness (αμεροληψία):**
  - When their expectation equals the parameter of estimation.
    $$\mathrm{E}\left(\hat{\theta}\right) = \theta$$
  - In general,
    - *bias* = $\mathrm{E}\left(\hat{\theta}\right) - \theta$
    - error of the estimator given a sample $\boldsymbol{x} = \{X_1, X_2, \ldots, X_n\}$
    $$e\left(\hat{\theta}(\boldsymbol{x})\right) = \hat{\theta}(\boldsymbol{x}) - \theta$$
- **Efficiency:**
  - We may have under consideration a set of **unbiased** estimators which all can estimate a specific parameter, θ. Among them we want the one whose estimates are closer to θ.
  - This is more or less equivalent to looking for the estimator which has a minimum variance between the variances of all the estimators within the set of estimators under consideration

5.2 (Good) Properties of Estimators (continued)

- Example of unbiased and efficient estimators
  - Let us return to our example of the estimation of the mean, for a sample of size *n* with values of interest denoted by $X_1, X_2, \ldots, X_n$, define the set (or class), Θ, of estimators of the form

$$\hat{\theta}_s = \frac{1}{k}\sum_{i=1}^{k} X_{n_i}, \qquad 1 \le n_1 < n_2 < \cdots < n_k \le n\,,\ 1 \le k \le n$$

    where the parameter *s* represents the specific selection among $X_1, X_2, \ldots, X_n$.
  - For every *s* we have

$$E(\hat{\theta}_s) = \frac{1}{k}E\left(\sum_{i=1}^{k} X_{n_i}\right) = \frac{1}{k}\sum_{i=1}^{k} E(X_{n_i}) = \frac{1}{k}k\mu = \mu$$

  - Therefore, all estimators in Θ are unbiased

5.2 (Good) Properties of Estimators (continued)

- Example of unbiased and efficient estimators (continued)
  - Which is the efficient estimator within this class?
  - Given that we have a random sample, are observations are independent, therefore the correlation between the X_i's is 0. We have

$$Var(\hat{\theta}_s) = \frac{1}{k^2}\sum_{i=1}^{k} Var(X_{n_i}) = \frac{1}{k^2}k\sigma^2 = \frac{\sigma^2}{k}$$

  - Therefore the estimator with the minimum variance is the one which corresponds to the largest k. Because the maximum value of k is n, the efficient estimator within Θ is

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Note that there is a lower bound beyond which it is not possible to approximate the true parameter given a sample of a specific finite size.

5.2 (Good) Properties of Estimators (continued)

- Consistency (συνέπεια)
  - An estimator is called consistent if it converges to the estimated parameter (in a probabilistic way) as the sample size increases to infinity.
  - Formally, let $\hat{\theta}$ be an estimator of θ and let us denote by $\hat{\theta}_n$ this estimator for a specific sample size, *n*. We say that $\hat{\theta}$ is a consistent estimator of θ if for any fixed $\varepsilon > 0$,

$$P\left(\left|\hat{\theta}_n - \theta\right| \geq \varepsilon\right) \to 0 \text{ as } n \to \infty$$

5.3 Probabilistic Convergence

- There are several types of convergence of random variables. We say that the sequence of r.v's X_n with corresponding c.d.f.s F_n, converges to the random variable X, with c.d.f. F,
  - *almost surely* (σχεδόν βέβαια) if $P\left(\lim_{n\to\infty} X_n = X\right) = 1$ (we denote it as $X_n \xrightarrow{a.s.} X$);
  - *in probability* (κατά πιθανότητα) if for any fixed $\varepsilon > 0$,
  $$P(|X_n - X| \geq \varepsilon) \to 0 \text{ as } n \to \infty$$
  (we denote it as $X_n \xrightarrow{P} X$);
  - *in the r-th mean* (σύγκλιση στον L^r), where $r \geq 1$, if
  $$\lim_{n\to\infty} E[|X_n - X|^r] = 0$$
  (we denote it as $X_n \xrightarrow{L^r} X$);
  - *in distribution* (κατά νόμο) if
  $$\lim_{n\to\infty} F_n(x) = F(x)$$
  in every continuity point, x, of F (we denote it as $X_n \xrightarrow{d} X$, or as $X_n \xrightarrow{L} X$).

5.3 Probabilistic Convergence (continued)

- **Remark:** In the previous definitions, *X* may be a constant (which can be considered as a trivial random variable).
- There is a hierarchy of convergences:
  - If $X_n \xrightarrow{P} X$ then $X_n \xrightarrow{d} X$
  - If $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{P} X$
  - If $X_n \xrightarrow{L^r} X$ then $X_n \xrightarrow{P} X$
  - However, there is no such a general relationship between a.s. and L^r convergences.

5.4 Important Theorems

- Law of large numbers (simple "weak" case) (WLLN)
  - Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent and identically distributed (iid) random variables with mean μ and standard deviation $\sigma < \infty$. Let also $\bar{X}_n$ be the average of the first *n* random variables (what we have previously denoted by $\hat{\mu}$). Then

$$\bar{X}_n \xrightarrow{P} \mu \text{ as } n \to \infty$$

  - **Remark1:** In simple cases we can prove the a.s. convergence of $\bar{X}_n$. We call these results: "strong" laws of large numbers.
  - **Remark2:** This theorem implies that when we have random samples from the same population, $\hat{\mu}$ is a consistent estimator of μ.
  - **Remark3:** The weak law of large numbers can be proved even in cases where the random variables are not independent. However, their dependence must be "weak".

5.4 Important Theorems (continued)

- The Central Limit Theorem (CLT)
  - Let $X_1, X_2, ..., X_n, \ldots$ be a sequence of independent and identically distributed (iid) random variables with mean μ and standard deviation $\sigma < \infty$ (again). Then
    $$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$
  - **Remark1:** The above result is equivalent to
    $$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{d} N(0,1)$$
    Note that the limiting distribution in the right hand side is invariant (does not depend on any of the parameters of the distribution of the X_i's)
  - **Remark2:** There are many variations of CLTs that also cover cases of "weakly" dependent random variables.