Prob. & Stats. - Week 8 Outline

Yi (Alice) Yin

November 5, 2017

1 Variance (continue)

1.1 Covariance

- **Definition:** The *covariance* of two random variables X, Y is

$$Cov(X,Y) = E\{[X - E(X)][(Y - E(Y)]\} = E(XY) - E(X)E(Y)$$

- **Properties**
  - Relation with Variance:

$$Cov(X,X) = Var(X)$$

  - "Change of units"

$$Cov(a_x X + b_x, a_y Y + b_y) = a_x a_y Cov(X,Y),$$

    where a_x, a_y, b_x, b_y are constants.

  - Variance of sums of random variables

$$Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$$

    More generally, for n random variables $\{X_i\}_{i=1}^n$,

$$Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} Cov(X_i, Y_j)$$

1.2 Uncorrelated random variables

- **Definition:** X, Y are uncorrelated if $Cov(X,Y) = 0$
- **Properties**
  - X, Y independent imply X, Y uncorrelated.
    The converse is *not* true: There exist uncorrelated random variables that are not independent.

- $Cov(X,Y) = 0 \iff Var(X+Y) = Var(X) + Var(Y)$
- More generally, if n random variables $\{X_i\}_{i=1}^n$ are pairwise uncorrelated, that is if $Cov(X_i, Y_j) = 0$ for every $i \neq j$, then

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

1.3 Covariance Coefficient

- **Motivation:** One would like a coefficient given an idea of how correlated two variables are. The correlation by itself is not suitable for two reasons:
  - The change of units formula implies that

    $$Cov(aX, aY) = a^2 Cov(X,Y),$$

    for any constant a. So the covariance can become arbitrarily large or small just by changing from meters to miilimiters or to kilometers .
  - $Cov(X,Y)$ has a unit which is the product of the units for X and Y, while the coefficient should be a number without any unit.
- **Definition** The **Covariance Coefficient of two random variables** X, Y is

$$\rho(x,y) = \begin{cases} \dfrac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} & \text{if } Var(X), Var(Y) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

- **Theorem**

$$\rho(X,Y) \in [-1, 1]$$

- **Observations**
  - $\rho(X,Y) = 1$ is interpreted as X, Y fully correlated. Examples of fully correlated random variables are $Y = X$ and, more generally, variables such that

    $$\frac{Y - \mu_Y}{\sigma_Y} = \frac{X - \mu_X}{\sigma_X}$$

  - $\rho(X,Y) = -1$ is interpreted as X, Y fully anticorrelated. Examples of fully anticorrelated random variables are $Y = -X$ and, more generally, variables such that

    $$\frac{Y - \mu_Y}{\sigma_Y} = -\frac{X - \mu_X}{\sigma_X}$$

2 Law of Large Numbers (LLN)

2.1 Independent and Identically Distributed (I.I.D.) R.V.s

- **Definition.** We say $\{X_i\}_{i=1}^n$ are **I.I.D.** if
  - All X_i follow the same law.
  - X_i are pairwise independent.

  for $\forall\ i \in \{1, 2, ..., n\}$.
- **Interpretation:**. The X_i are interpreted as independent repetitions of the same experiment. If $X$is the random variable describing this experiment, then all X_i are independent and have the same law as X. This last fact is often denoted by $X_i \sim X$.
- **Observations:** In particular, IID random variables
  - have the same mean, $E(X_i) = \mu$,
  - have the same variance, $Var(X_i) = \sigma^2$ and
  - are uncorrelated.

2.2 Averages of random variables

- **Definition.** The **average** or **empirical mean** of a set of r.v.s $\{X_i\}_{i=1}^n$ is

$$\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

- **Properties:**
  - If $E(X_i) = \mu$, then

$$E(\overline{X}_n) = \mu\ .$$

  - If $Var(X_i) = \sigma^2$ and the variables are *uncorrelated*, then

$$Var(\overline{X}_n) = \frac{\sigma^2}{n}\ .$$

  - The two preceding properties hold, in particular, if the variables X_i are IID.

2.3 Chebyshev Inequality

- If X is a random variable with $E(X) = \mu$ and $Var(X) = \sigma^2$, then

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

- If $\{X_i\}_{i=1}^n$ are I.I.D with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, then

$$P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n}\ . \tag{1}$$

2.4 Weak LLN

- **Theorem.** If $\{X_i\}_{i=1}^n$ are I.I.D. with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, then

$$P(|\bar{X}_n - \mu| > \epsilon) \xrightarrow[n\to\infty]{} 0, \ \forall \epsilon > 0$$

- **Comments:**
  - Inequality (1) tells us how fast this probability converges to 0. This can be used to determine the number of repetitions needed to estimate μ with high probability within a prescribed error margin.
  - In fact, the theorem only requires that the random variables X_i have equal mean and variance and that they be uncorrelated. Full independence is not needed.
  - As stated, the theorem does *not* hold for the following distributions:
    - Cauchy (the mean is not defined),
    - $Par(\alpha \leq 1)$ (the mean is not finite), and
    - $Par(1 < \alpha \leq 2)$ (the variance is not finite).

2.5 Strong LLN

- **Theorem.** If $\{X_i\}_{i=1}^n$ are I.I.D. with $E(X_i) = \mu$, then

$$P\big(\{\lim_{n\to\infty} \overline{X}_n = \mu\}\big) = 1$$

- **Comments:**
  - This version of the LLN does require full independence, but does not require a finite variance. It therefore applies to $Par(1 < \alpha \leq 2)$.
  - The strong LLN means that averages of *every* sequence of outcomes of a repeated experiment lead to the actual mean μ.
  - It can be seen that, if the variables have finite variance, the strong law implies the weak law (that is what is meant when labelling this version of the LLN as strong).

2.6 Empirical frequencies

The empirical frequencies allow to use the LLN to determine all features of the law of the common variable X.

- **Motivation**
  - Objective: to determine $P(X \in C)$
  - Trick: To use the LLN, define Y s.t. $E(Y) = P(X \in C)$

- This is done through the **Indicator Function** of $\{X \in C\}$

$$Y(\omega) = \begin{cases} 1 & \text{if } X(\omega) \in C \\ 0 & \text{otherwiese} \end{cases}$$

- We have:
    * $E(Y) = P(X \in C)$
    * $Var(Y) = P(X \in C)\, P(X \notin C)$

- **Definition.** The *empirical frequency* of an event C is the empirical average of the corresponding indication function, namely

$$\overline{Y}_n(C) \;=\; \frac{\sum_{i=1}^{n} Y_i}{n} \;=\; \frac{\#\text{ of times in C}}{n}$$

2.7 LLN for frequencies

- **Weak LLN:** If the X_i variables are IID, then

$$\begin{aligned} P\big(\big|\overline{Y}_n - P(X \in C)\big| > \epsilon\big) \quad &\le \quad \frac{P(X \in C)\, P(X \notin C)}{n\epsilon^2} \\ &\le \quad \frac{1}{n\epsilon^2} \\ &\xrightarrow[n\to\infty]{} \quad 0 \end{aligned}$$

  for all $\epsilon > 0$.

- **Strong LLN:** If the X_i variables are IID, then

$$P\big(\big\{\lim_{n\to\infty} \overline{Y}_n = P(X \in C)\big\}\big) \;=\; 1$$

  for all $\epsilon > 0$.

- **Observations**
    - Both forms of the LLN are valid for **all** X, regardless on whether it has well defined mean and variance. Indeed, in all cases Y has well defined mean $[= P(X \in C)]$ and variance $[= P(X \in C)P(X \notin C)]$.
    - The weak LLN can be used to estimate the proximity of empirical frequencies to the actual probabilities $P(X \in C)$.
    - The strong LLN implies that frequencies measured from *every* sequence of outcomes of a repeated experiment leads to the actual probability.