Statistics

Moments

Shiu-Sheng Chen

Department of Economics
National Taiwan University

Fall 2019

Section 1

Moments

Moments

- Moments can help us to summarize the distribution of a random variable.
  - Analogy to: the height, weight, hair color...etc. of a person
  - Essential but Concise
- Two important moments:
  - Expectation
  - Variance

Expectation

Definition (Expectation)

Let $S = \text{supp}(X)$. The expectation of a random variable X is defined as

$$E(X) = \begin{cases} \sum_{x \in S} x f(x) & \text{discrete} \\ \int_{x \in S} x f(x) dx & \text{continuous} \end{cases}$$

- Expected value; Mean (value)
- A probability-weighted sum of the possible values.
  - Expectation is a constant.
  - Conventional notation: $E(X) = \mu$

Example: Fair Price for a Stock

- An investor is considering whether or not to invest in a stock for one year.
- Let Y represent the amount by which the price changes over the year with the following distribution

y	-2	0	1	4
$f(y)$	0.1	0.4	0.3	0.2

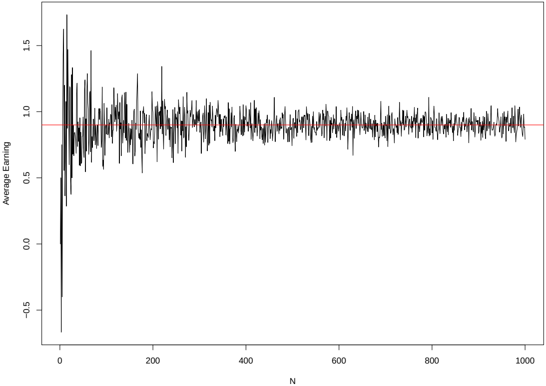
- Then the expected earning is

$$E(Y) = 0.9$$

That is, "on average, the investor expects to earn 0.9." $\Leftarrow$ What does this mean?

Simulation Results and Interpretation

- If you invest in this stock N years. Let Y_i denote the price change for year i, and the average earning is thus $\frac{\sum_{i=1}^{N} Y_i}{N}$
- We can see that $\frac{\sum_{i=1}^{N} Y_i}{N}$ is very close to $E(Y) = 0.9$ when N is large.



Expectation

Theorem (The Rule of Lazy Statistician)

Let X be a random variable, and let $g(\cdot)$ be a real-value function. Then

$$E(g(X)) = \begin{cases} \sum_x g(x)f(x) \\ \int_x g(x)f(x)dx \end{cases}$$

- Example:

$$X = \begin{cases} 2 & \text{with } P(X=2) = 1/3 \\ 1 & \text{with } P(X=1) = 1/3 \\ -1 & \text{with } P(X=-1) = 1/3 \end{cases}$$

  - Consider $g(X) = X^2$, find $E(g(X)) = ?$

Variance and Standard Deviation

Definition (Variance/SD)

The variance of a discrete random variable X is defined by

$$Var(X) = E\left[(X - E(X))^2\right] = \begin{cases} \sum_x (x - E(X))^2 f(x) \\ \int_x (x - E(X))^2 f(x)dx \end{cases}$$

- It describes how far values lie from the mean.
- Conventional notation: $Var(X) = \sigma^2$
- The standard deviation is $SD(X) = \sqrt{Var(X)}$, and denoted by $SD(X) = \sigma$.

Constant as a Random Variable

- Given a constant c, then

$$E(c) = c,$$

and

$$Var(c) = 0.$$

- Therefore,

$$E(E(X)) = E(X)$$

Some Important Properties

- Given constants a and b:

$$E(aX + b) = aE(X) + b$$

$$Var(aX + b) = a^2 Var(X)$$

- It can shown that

$$E(X - E(X)) = 0$$

$$Var(X) = E(X^2) - [E(X)]^2$$

Example: Fair Price for a Stock

Definition

The fair price of a stock is defined by a price such that the expected return equals the risk free rate.

- Suppose that the stock price is p.
- The return is
$$\frac{(p+Y)-p}{p}=\frac{Y}{p}$$
- As an alternative, the investor can put the money in the bank with a 5% interest rate (risk-free).
- Recall that $E(Y)=0.9$. Hence, $E\left(\frac{Y}{p}\right)=0.05$ shows that $p=18$ is the fair price of the stock.

Expectation as the Best Constant Predictor

- Consider a constant predictor of X, say c.
- Mean Square Prediction Error

$$MSPE = E\left[(X-c)^2\right]$$

- It can be shown that

$$E(X) = \arg\min_c E\left[(X-c)^2\right]$$

More on Expectation

- In general, unless $g(\cdot)$ is linear,

$$E(g(X)) \neq g(E(X))$$

  - For instance, in the previous example, $g(X) = X^2$,

$$E(X^2) = 2 \neq \frac{4}{9} = [E(X)]^2$$

- One more example,

$$E\left(\frac{1}{X}\right) \neq \frac{1}{E(X)}$$

- Proof: by Jensen's Inequality

Jensen's Inequality

Theorem

If X is a random variable and $g(\cdot)$ is a convex function, then

$$E(g(X)) \geq g(E(X))$$

Proof.

Since $g(\cdot)$ is a convex function, there exist some constants a and b such that $g(X) \geq aX + b$, and $g(E(X)) = aE(X) + b$. $\square$

Standardized Random Variables

- As expectation and variance are the two most important moments, sometimes we will denote the random variable as

$$X \sim (E(X), Var(X)) \quad \text{or} \quad X \sim (\mu, \sigma^2)$$

Definition (Standardized Random Variables)

Given $X \sim (\mu, \sigma^2)$, and let

$$Z = \frac{X - \mu}{\sigma}$$

Then $Z \sim (0, 1)$ is called a standardized random variable.

Moments

- k-th Moments

$$E(X^k)$$

- k-th Central Moments

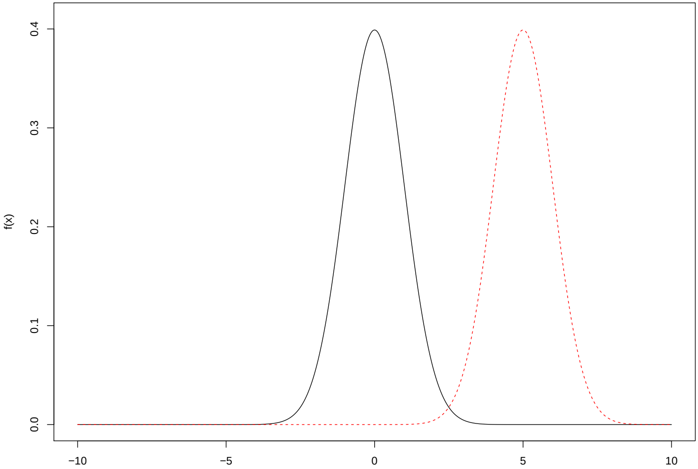
$$E[(X - E(X))^k]$$

- k-th Standardized Moments

$$\gamma_k = E\left(\left[\frac{X - E(X)}{\sqrt{Var(X)}}\right]^k\right)$$

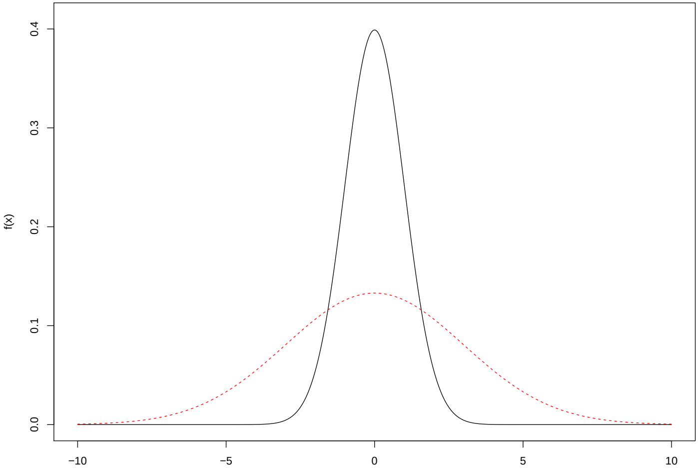
Expectation

- $E(X) = 0$ vs. $E(X) = 5$ $(Var(X) = 1)$



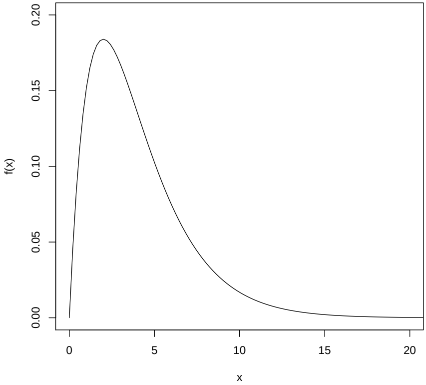
Variance

- $Var(X) = 1$ vs. $Var(X) = 9$ $(E(X) = 0)$



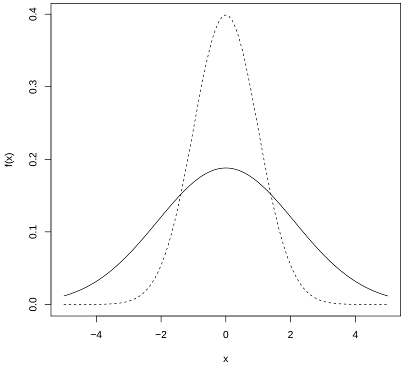
Skewness

- 3rd standardized moment (Skewness): $\gamma_3 = E\left[\left(\frac{X-E(X)}{\sqrt{Var(X)}}\right)^3\right]$
- $\gamma_3 > 0$



Kurtosis

- 4th standardized moment (Kurtosis): $\gamma_4 = E\left[\left(\frac{X-E(X)}{\sqrt{Var(X)}}\right)^4\right]$
  - Excess Kurtosis $= \gamma_4 - 3$
  - Fat tail: Excess Kurtosis > 0



Section 2

Moment Generating Functions

Moment Generating Functions

Definition (MGF)

Let X be a discrete (continuous) random variable, and the pmf (pdf) is $f(x)$. Given $h > 0$ and for all $-h < t < h$, if the following function

$$M_X(t) = E(e^{tX})$$

exists and is finite, it is called the moment generating function (MGF) of the random variable X.

- One use the MGFs is that, in fact, it can generate moments of a random variable.

Properties

Theorem (Moment Generating)

$$E(X^k) = M_X^{(k)}(0) = M_X^{(k)}(t)\big|_{t=0},$$

where $M_X^{(k)}(t)$ *denotes the* k*-th derivative of* $M_X(t)$*.*

- Proof: expand e^{tX} as

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \frac{(tX)^4}{4!} + \cdots$$

Properties

Theorem (Uniqueness)

For all $t \in (-h, h)$, if $M_X(t) = M_Y(t)$, then X and Y has exactly the same distribution, $F_X(c) = F_Y(c)$ for all $c \in \mathbb{R}$.

- Proof: Beyond the scope of this course (via so-called the inverse Fourier transform).

Properties

Theorem (MGF of Linear Transformations)

Given the MGF of X is $M_X(t)$. Let $Y = aX + b$, then

$$M_Y(t) = e^{bt} M_X(at).$$

- Proof: By definition.

Examples

Find the MGFs of the following random variables:

- $X \sim \text{Bernoulli}(p)$
- $X \sim \text{Binomial}(n, p)$
- $X \sim \text{Uniform}[l, h]$

Section 3

Covariance and Correlation

Expected Values of Functions of Bivariate Random Variables

Definition

Let X and Y be discrete (continuous) random variables with joint pmf (pdf) $f_{XY}(x,y)$. Let $g(X,Y)$ be a function of these two random variables, then:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) f_{XY}(x,y)$$

$$E[g(X,Y)] = \int_x \int_y g(x,y) f_{XY}(x,y) dy dx$$

Covariance

Definition (Covariance)

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

- It is typically denoted by σ_{XY}.
- A measurement of comovement among two random variables.

$x - E(X)$	$y - E(Y)$	$Cov(X,Y)$
$+$	$+$	$+$
$-$	$-$	$+$
$+$	$-$	$-$
$-$	$+$	$-$

- It can be shown that

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Properties

Theorem

Given constants a, b, c, and d

- $E(aX+bY) = aE(X) + bE(Y)$
- $Cov(X,X) = Var(X)$
- $Cov(X,c) = 0$
- $Var(aX+bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$
- $Cov(X,Y) = E(XY) - E(X)E(Y)$

Correlation Coefficient

Definition (Correlation Coefficient)

The correlation coefficient is defined by

$$\rho_{XY} = Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

- A unit-free measure of comovement.

Correlation Coefficient

Theorem

The correlation coefficient lies between 1 and -1:

$$-1 \leq \rho_{XY} \leq 1$$

- Proof: by Cauchy-Schwarz inequality,

$$[E(UV)]^2 \leq E(U^2)E(V^2)$$

Correlation Coefficient

- Note:
  - $\rho_{XY} = 1$ (perfect correlation)
  - $\rho_{XY} = -1$ (perfect negative correlation)
  - $\rho_{XY} = 0$ (zero correlation, no correlation, uncorrelated)
- However, no correlation does not mean that there is no relationship between X and Y
- It just suggests that there is no linear relationship between X and Y

Section 4

Independent Bivariate Random Variables

Expectation of Functions of Independent Bivariate Random Variables

Theorem

Let X and Y are independent variables. Then

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Independent Bivariate Random Variables and MGF

Theorem

X and Y are independent random variables. Their MGFs are $M_X(t)$ and $M_Y(t)$, respectively. Let $Z = X + Y$, then

$$M_Z(t) = M_X(t)M_Y(t)$$

- Proof. By definition and the previous theorem.
- Example: Revisit the MGF of a Binomial(n,p) random variable

Independent Bivariate Random Variables

Theorem

Given that X and Y are independent:

- $E(XY) = E(X)E(Y)$
- $Cov(X, Y) = 0$
- $Var(X + Y) = Var(X) + Var(Y)$

Independent vs. Uncorrelated

- X, Y independent implies X, Y uncorrelated, however, the reverse is not true.
  - Independence require all possible realizations x and y to satisfy

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

  - To check X, Y uncorrelated, only one equation needs to hold:

$$\sum_x \sum_y (x - E(X))(y - E(Y))P(X = x, Y = y) = 0.$$

Example

- Consider random variable X has the following distribution:

x	$P(X = x)$
-1	1/3
0	1/3
1	1/3

- Now let $Y = X^2$
- It can be shown that $Cov(X, Y) = 0$ but clearly they are not independent.

Section 5

Conditional Expectation and Variance

Conditional Expectation

Definition

The conditional expectation of Y given $X = x$ is

$$E(Y|X = x) = \sum_y y f_{Y|X=x}(y)$$

$$E(Y|X = x) = \int_y y f_{Y|X=x}(y) dy$$

- Hence, $E(Y|X = x) = g(x)$
- Since $E(Y|X = x)$ is a function of x, it follows that

$$E(Y|X) = g(X)$$

Conditional Variance

Definition

The conditional variance of Y given $X = x$ is

$$Var(Y|X = x) = E([Y - E(Y|X = x)]^2|X = x)$$

- Hence, $Var(Y|X = x) = h(x)$
- Since $Var(Y|X = x)$ is also a function of x, it follows that

$$Var(Y|X) = h(X)$$

- It can be shown that (will be shown later)

$$Var(Y|X = x) = E(Y^2|X = x) - [E(Y|X = x)]^2$$

Example

- Given two continuous random variables X and Y with joint pdf

$$f_{XY}(x, y) = \frac{3}{2},$$

$$\text{supp}(Y) = \{y | x^2 < y < 1\}, \ \text{supp}(X) = \{x | 0 < x < 1\}$$

- Find $f_{Y|X=x}(y)$, $E(Y|X = x)$, and $E(Y|X = 1/2)$.

Important Theorems

Theorem

- *Useful Rule*

$$E[h(X)Y|X] = h(X)E[Y|X]$$

- *Simple Law of Iterated Expectation*

$$E(E[Y|X]) = E(Y)$$

$$E(E[XY|X]) = E(XY)$$

- Application:

$$Var(Y|X) = E(Y^2|X) - [E(Y|X)]^2$$

Example

- Given two continuous random variables X and Y with joint pdf

$$f_{XY}(x,y) = \frac{3}{2},$$

$$\mathrm{supp}(Y) = \{y|x^2 < y < 1\},\ \mathrm{supp}(X) = \{x|0 < x < 1\}$$

- Find $E(Y^2|X=x)$, $Var(Y|X=x)$, and $Var(Y|X=1/2)$.

Important Theorems

Theorem (Variance Decomposition)

$$Var(Y) = Var\Big(E(Y|X)\Big) + E\Big(Var(Y|X)\Big)$$

- Example:

$$X \sim \text{Bernoulli}(P)$$

where

$$P \sim \text{Uniform}[0,1]$$

Find $E(X)$ and $Var(X)$.

Important Theorems

Theorem (Best Conditional Predictor)

Conditional expectation $E(Y|X)$ is the best conditional predictor of Y in the sense of minimizing the conditional mean squared error:

$$E(Y|X) = \arg\min_{g(X)} E[(Y - g(X))^2]$$

Example: GPA vs. Study Hours

- Let Y = GPA, X = Study Hours
- We would like to know $E(Y|X)$ (to forecast Y)
- We further assume that $E(Y|X)$ is a linear function:

$$E(Y|X) = \alpha + \beta X$$

- It can be shown that

$$\beta = \frac{E(XY) - E(X)E(Y)}{E(X^2) - E(X)^2} = \frac{Cov(X,Y)}{Var(X)}$$

$$\alpha = E(Y) - \beta E(X)$$

Example: GPA vs. Study Hours

- We can define the forecast error as

$$\epsilon \equiv Y - E(Y|X) = Y - (\alpha + \beta X)$$

- Hence,

$$Y = \alpha + \beta X + \epsilon$$

  - Interpretation: your GPA is determined by
    (a) Systematic Part: $\alpha + \beta X$, which can be explained by study hours
    (b) Irregular Part: ϵ, which captures other factors other than study hours. For instance, good/bad luck, mood, illness, etc.

Section 6

Multivariate Random Variables

Expected Values of Functions of Random Variables

- For discrete random variables, the expected values of $g(X_1, X_2, \ldots, X_n)$ is given by

$$E[g(X_1, X_2, \ldots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} g(x_1, x_2, \ldots, x_n) f_X(x_1, x_2, \ldots, x_n)$$

- For continuous random variables,

$$E[g(X_1, X_2, \ldots, X_n)] = \int_{x_1} \cdots \int_{x_n} g(x_1, \ldots, x_n) f_X(x_1, \ldots, x_n) dx_n \cdots dx_1$$

Properties

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} Cov(X_i, X_j)$$

Expectation of Functions of Independent Random Variables

Theorem

Let $X_1, X_2, \ldots, X_n$ are independent variables. Then

$$E[h(X_1)h(X_2)\cdots h(X_n)] = E[h(X_1)]E[h(X_2)]\cdots E[h(X_n)]$$

Independent Random Variables and MGF

Theorem

X_1, X_2,. . . , X_n are independent with MGF: $M_{X_1}(t)$, $M_{X_2}(t)$,. . ., $M_{X_n}(t)$. Let $Y = \sum_{i=1}^{n} X_i$, then

$$M_Y(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

IID Random Variables

- Given that $\{X_i\}_{i=1}^n$ are i.i.d. random variables.
- Clearly,

$$E(X_1) = E(X_2) = \cdots = E(X_n)$$

$$Var(X_1) = Var(X_2) = \cdots = Var(X_n)$$

$$Cov(X_i, X_j) = 0 \quad \text{for any } i \neq j$$

- I.I.D. random variables with mean μ and variance σ^2 are denoted by

$$\{X_i\}_{i=1}^n \sim^{i.i.d.} (\mu, \sigma^2)$$

Properties of i.i.d. Random Variables

Theorem

Let $\{X_i\}_{i=1}^n$ *are i.i.d. random variables with* $E(X_i) = \mu$ *and* $Var(X_i) = \sigma^2$*, and let*

$$Y = \sum_{i=1}^{n} X_i.$$

Then

$$E(Y) = n\mu,$$

$$Var(Y) = n\sigma^2.$$

Example

- Let

$$\{X_i\}_{i=1}^n \sim^{i.i.d.} \text{Bernoulli}(p)$$

  - That is,

$$E(X_i) = p \quad \text{and} \quad Var(X_i) = p(1-p)$$

- Let

$$Y = \sum_{i=1}^n X_i$$

  - What is the distribution of Y?
  - Find $E(Y)$ and $Var(Y)$