Part 2

Statistics Revision

EE325

Introductory Econometrics

Revision Aug 2020

Event, sample space, probability

Let A be an event of interest, occurring within a given sample space S and $P(A)$ be the probability that A will occur, $P(A)$ is defined as

- $P(A) = \frac{number\ of\ times\ event\ A\ will\ occur}{number\ of\ all\ possible\ outcome\ in\ sample\ space}$

 # Example: tossing 2 fair coins, the sample space is

- $S = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

If the event of interest is having at least a coin turning head (H) is

- $A = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}\}$.

The probability of this event is then

- $P(A) = \frac{3}{4}$

Probability Axioms

(1) $0 \leq P(A) \leq 1$

(2) $P(S) = 1$

(3) If and A are B mutually exclusice, then $P(A \cup B) = P(A) + P(B)$

Random variable

Let X be the results of an experiment in the form of value, which value is given by one of the result, is called X a **random variable**.

 # Example: tossing 2 fair coins again, let X be 0 if the result shows at least a coin turned up head, be 1 otherwise. The sample space was defined as

- $S = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

Transforming theses events into random variable, we get

- $S_X = \{0,1\}$

But we know that the probability of 0 and 1 are different since 1 is more likely to occur. Therefore, if we put probability function with specific value of random variable X, we have

- $P(X = 0) = \frac{3}{4}$
- $P(X = 1) = \frac{1}{4}$

<u>**Types of random variable**</u>

(1) Discreate random variable is a random variable that can take specific values of event. E.g. the coin example.

(2) Continuous random variable is a random variable that can takes infinite amount of values of event. E.g. Height of BE students.

Probability density function (PDF)

A PDF is a function whose value at any given sample (or point) in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample.

 # Example: Let X be a random variable of total points from rolling 2 fair dices, the sample space would be

- $S_X = \{2,3,4,5,6,7,8,9,10,11,12\}$

PDF is defined as

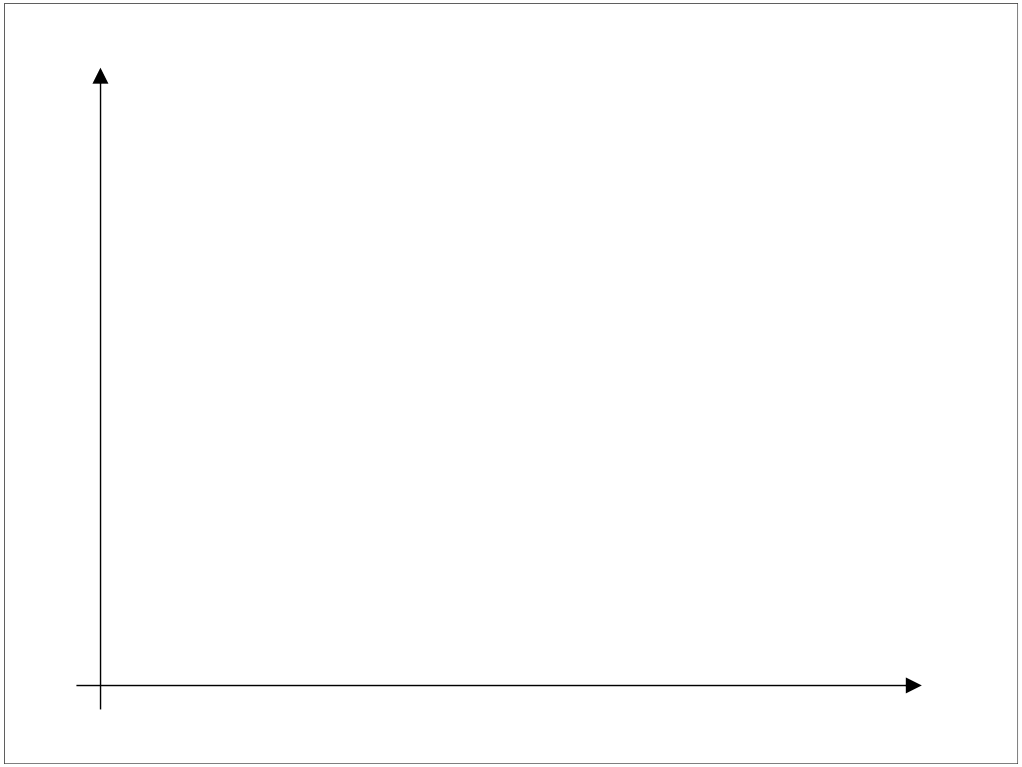
- $f(x_i) = P(X = x_i)$ for $x_i \in S_X$
- $f(x_i) = 0$ for $x_i \notin S_X$

Let's draw a PDF of this scenario.

Now figure out these probability.

- $P(X = 4) =$
- $P(X = 7) =$
- $P(X < 3) =$
- $P(X \leq 4) + P(X > 9) =$

PDF of rolling 2 fair dices



Properties of PDF

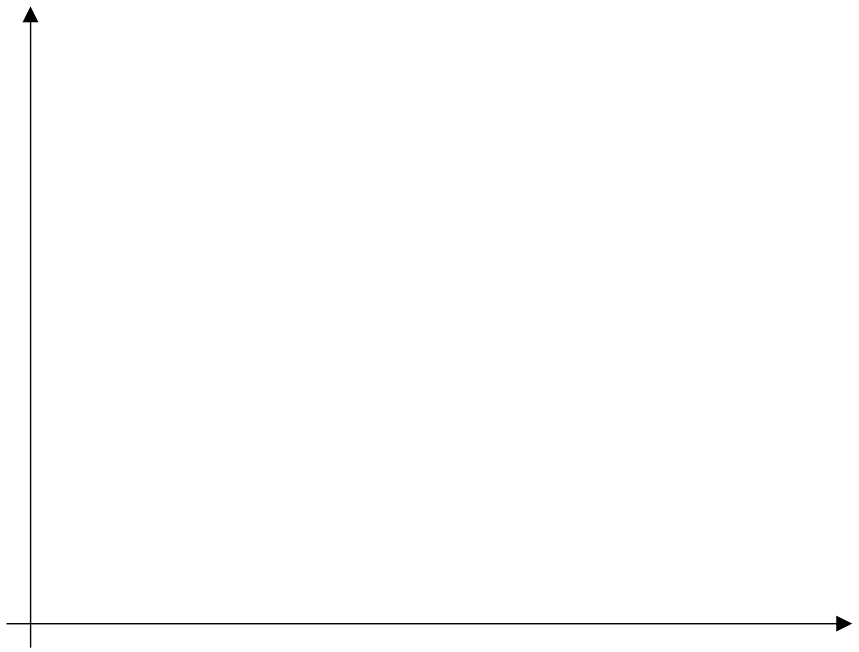
(1) Discreate PDF

- $0 \leq f(x) \leq 1$
- $\sum_{-\infty}^{\infty} f(x) = 1$
- $\sum_{a}^{b} f(x) = P(a \leq X \leq b)$

(2) Continuous PDF

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $\int_{a}^{b} f(x)dx = P(a \leq X \leq b)$

An example of continuous PDF



Bivariate probability density function

(1) Joint probability density function

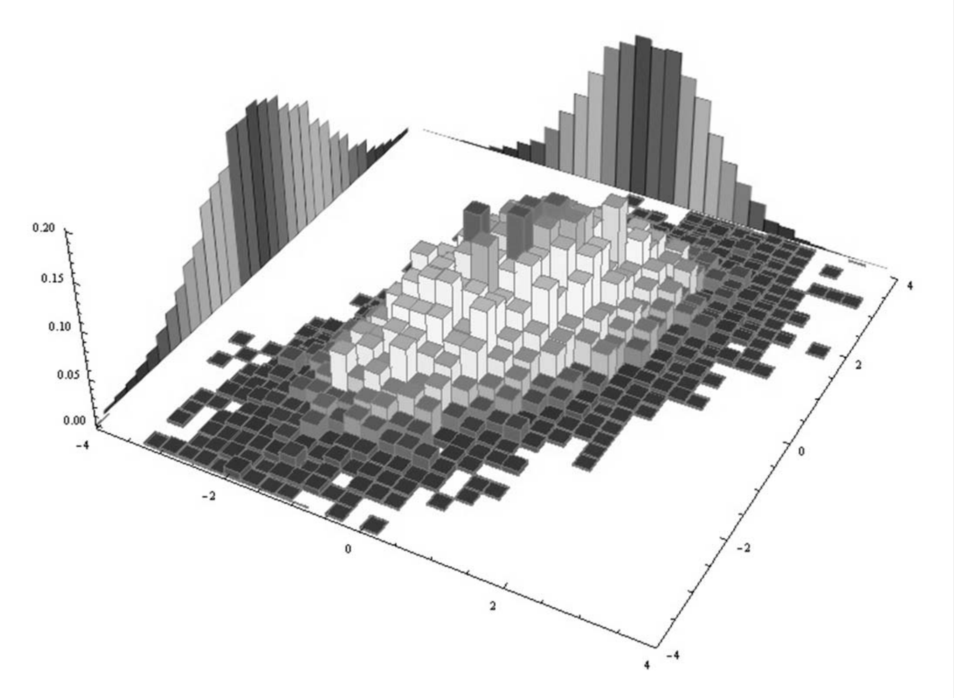
Let X and Y be two random variables, the joint probability distribution for X and Y is a probability distribution that gives the probability that each of X and Y falls in any particular range or discrete set of values specified for that variable. Joint PDF is defined as

- $f(x,y) = P(X = x, Y = y)$

 # Example: two archers are competing shooting a target for 3 times each. Let X and Y be number of times archer 1 and archer and archer 2 hit the target respectively. Thousands of round has been competed and the probability is computed as a result in this table.

		X		
		1	2	3
Y	1	0.35	0.2	0.07
	2	0.1	0.05	0.09
	3	0.08	0.04	0.02

Joint PDF



- Find the probability of archer 1 hitting 2 times while archer 2 hitting 3 times straight.

- Find $P(X = 1, Y \geq 2)$.

Bivariate probability density function

(2) Marginal probability

Marginal probability, with respect for joint PDF of X and Y, is the probability of a single event occurring, independent of other events.

Marginal probability of X and Y is respectively defined, for discrete random variable, as

- Marginal probability of X is $\sum_y f(x, y)$
- Marginal probability of Y is $\sum_x f(x, y)$

and for continuous variable are

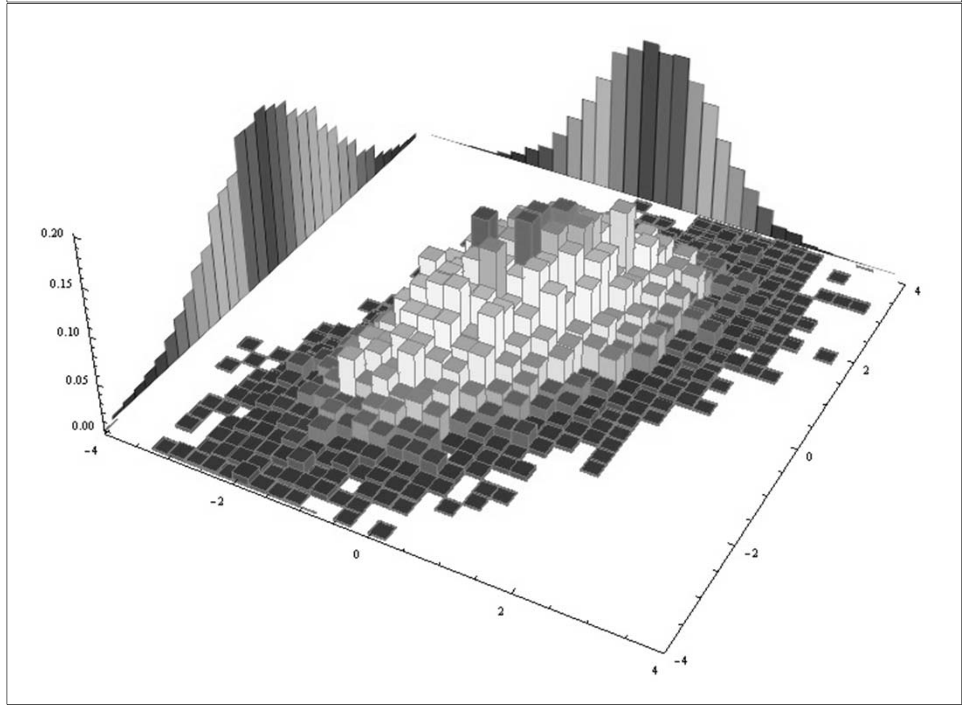
- Marginal probability of X is $\int_{-\infty}^{\infty} f(x, y)dy$
- Marginal probability of Y is $\int_{-\infty}^{\infty} f(x, y)dx$

 # Example: using the same archer example,

- Find marginal probability of X when $X = 1$

- Find marginal probability of Y when $Y = 3$

Visualize marginal probability



Bivariate probability density function

(3) Conditional probability

Conditional probability is a measure of the probability of an event occurring given that another event has (by assumption, presumption, assertion or evidence) occurred. Conditional probability is defined as

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

From the definition, we can figure the conditional probability by using joint PDF. Using the same example,

		X		
		1	2	3
Y	1	0.35	0.2	0.07
	2	0.1	0.05	0.09
	3	0.08	0.04	0.02

- Find $f(X = 1|Y = 2) =$

- Find $f(Y = 2|X = 3) =$

Bivariate probability density function

(3) Conditional probability

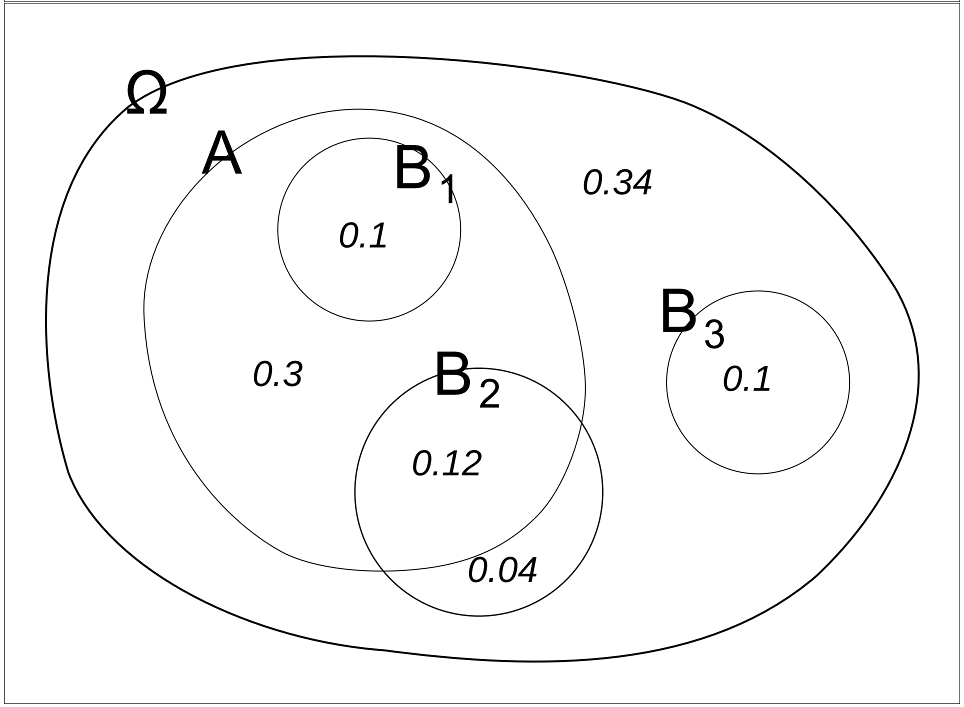
Here is another good visualization.

- Find $f(A|B_1) =$

- Find $f(A|B_2) =$

- Find $f(A|B_3) =$

Conditional probability



Bivariate probability density function

(4) Statistical independence

Two random variables are considered independent if and only if the condition below is satisfied.

- $f(x, y) = f(x) \cdot f(y)$

 # Example: with the archer example, if two archers are shooting one at a time, archer 1 shoots first followed by archer 2. Archer 2 may find it more stressful if archer 1 can hit the target. However, if they shoot the target independently, one's performance may not be affected by another's.

 # Example: prove that and are independent.

- $f(x, y) = f(x) \cdot f(y)$

		X		
		1	2	3
Y	1	1/9	1/9	1/9
	2	1/9	1/9	1/9
	3	1/9	1/9	1/9

Expectation, variance, covariance, correlation

(1) Expected value

Expected value of a random variable is a generalization of the weighted average and intuitively is the arithmetic mean of independent realizations of that variable. Expected value is defined as

- For discrete random variable: $E(X) = \sum_{i=1}^{n} x_i \cdot f(x_i)$
- For continuous random variable: $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$

<u>**Properties of expected value**</u>

- $E(a) = a$ for any constant a
- $E(aX) = aE(X)$
- $E(aX + b) = aE(X) + b$
- $E(X \pm Y) = E(X) \pm E(Y)$
- $E(XY) = E(X) \cdot E(Y)$ if and only if X and Y are independent.

 # Example: Paope is trying hard in her econometrics class. She knows that her probability of getting grades are.

Grade	Prob
A	0.3
B	0.4
C	0.15
D	0.1
F	0.05

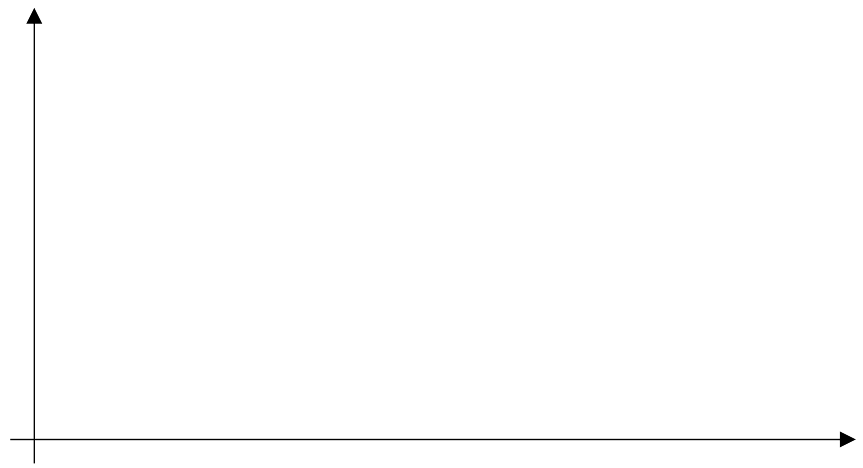
- Find the expected value of her grade $E(X) = \sum_{i=1}^{n} x_i \cdot f(x_i)$

- Plot this PDF and locate the expected value.

Expectation, variance, covariance, correlation

(1) Expected value

Paope and her struggle in econometrics class



 # Example: find the expected value of this PDF

- $f(X) = \frac{1}{9}x^2$ for $0 \leq x \leq 3$

Expectation, variance, covariance, correlation

(2) Conditional expectation

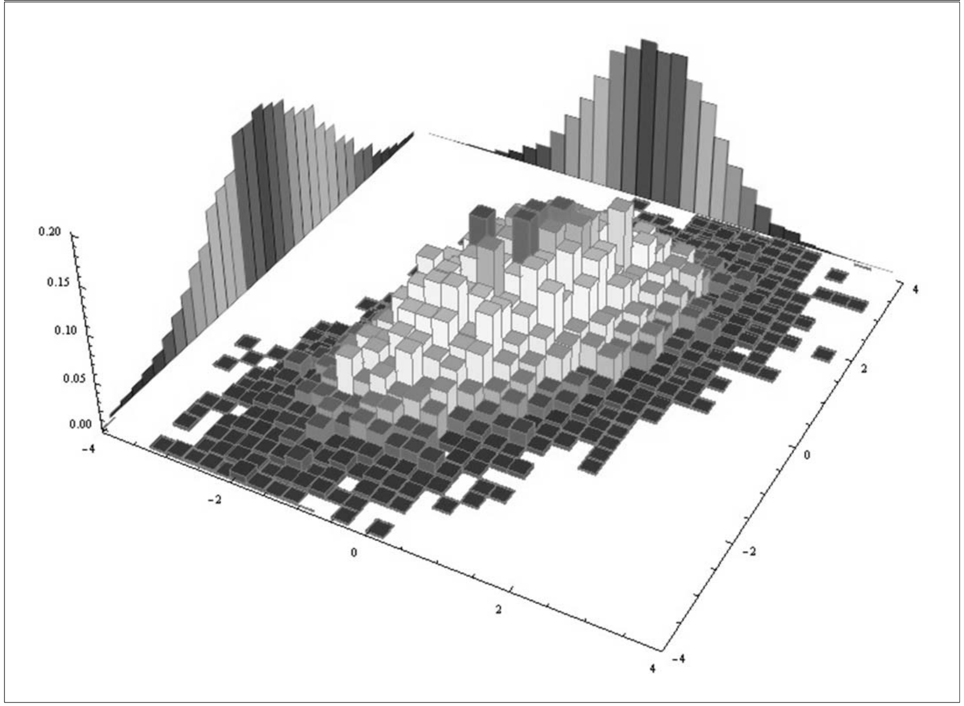
Let $f(X,Y)$ be a joint probability density function, the expectation of X conditional on some value of Y is

- For discrete random variable: $E(X|Y) = \sum_X x_i \cdot f(X|Y=y)$
- For continuous random variable: $E(X|Y) = \int_{-\infty}^{\infty} x \cdot f(X|Y=y)dx$

 # Example: find $E(X|Y=3)$ from the PDF below.

		X			
		-2	0	2	3
Y	3	0.27	0.08	0.16	0
	6	0	0.04	0.1	0.35

Bring back the illustration



Expectation, variance, covariance, correlation

(3) Variance

Variance is a measure of data dispersion from the expected value. Given that μ is the expected value of X, then

- For discrete random variable: $Var(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot f(x_i)$
- For continuous random variable: $Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx$

Proof of variance

We can prove that $Var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$

Properties of variance

- $Var(a) = 0$ for any constant a
- $Var(aX + b) = a^2 Var(X)$
- $Var(X \pm Y) = Var(X) \pm Var(Y)$ if and only if X and Y are independent.

 # Example: find the variance from the given PDF.

X	-2	1	2
$f(X)$	5/8	1/8	2/8

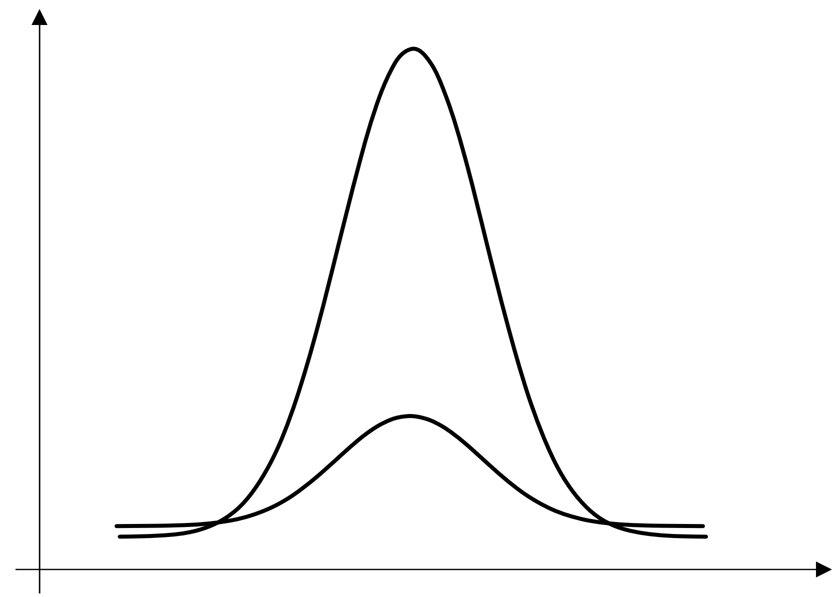
Expectation, variance, covariance, correlation

(3) Variance

 # Example: find the variance from the given PDF.

- $f(X) = \frac{1}{9}x^2$ for $0 \leq x \leq 3$

Comparing high and low variance distribution



Expectation, variance, covariance, correlation

(4) Conditional variance

Conditional variance is a measure variance, but coupled with a condition on another variable, defined as

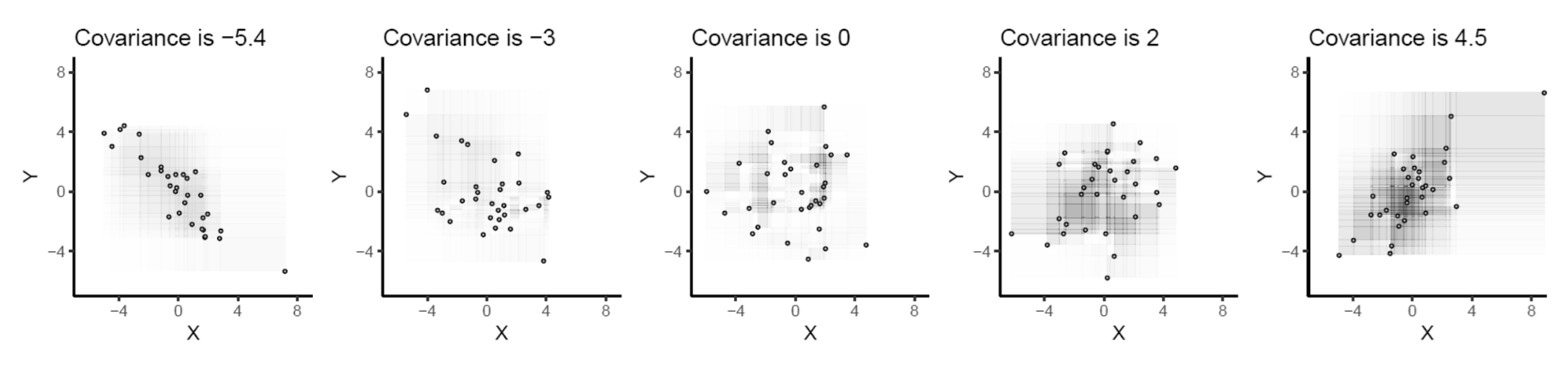
- For discrete random variable: $Var(X|Y) = \sum_X [x_i - E(X|Y=y)]^2 \cdot f(X|Y=y)$
- For continuous random variable: $Var(X|Y) = \int_{-\infty}^{\infty} [x - E(X|Y=y)]^2 \cdot f(X|Y=y)dx$

(5) Covariance

Let X and Y be two random variables with expected value of μ_X and μ_Y respectively, the covariance is a measure of the joint variability of two random variables. If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values. Defined as

- For discrete random variable: $cov(X,Y) = E\{(X-\mu_X)(Y-\mu_Y)\} = E(XY) - \mu_X\mu_Y$
- For continuous random variable: $cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X-\mu_X)(Y-\mu_X) \cdot f(x,y)dxdy - \mu_X\mu_Y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XYf(x,y)\,dxdy - \mu_X\mu_Y$

Comparing covariance



Expectation, variance, covariance, correlation

(5) Covariance

Problems with interpretation

"A large covariance can mean a strong relationship between variables. However, you can't compare variances over data sets with different scales (like pounds and inches). A weak covariance in one data set may be a strong one in a different data set with different scales.

The main problem with interpretation is that the wide range of results that it takes on makes it hard to interpret. For example, your data set could return a value of 3, or 3,000. This wide range of values is cause by a simple fact; The larger the X and Y values, the larger the covariance."

Further properties of variance

If X *and* Y are ***not*** independent, then

- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2cov(X, Y)$

Properties of covariance

1. If X *and* Y are independent, the covariance is zero.

Proof:

2. If $cov(a + bX, c + dY) = bd \cdot cov(X, Y)$

Proof:

Expectation, variance, covariance, correlation

(6) Correlation coefficient

Correlation coefficient is a numerical measure of some type of correlation, meaning a statistical relationship between two variables, denoted by ρ_{XY} or $\mathrm{corr}(X, Y)$.

- $\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \epsilon[-1,1]$

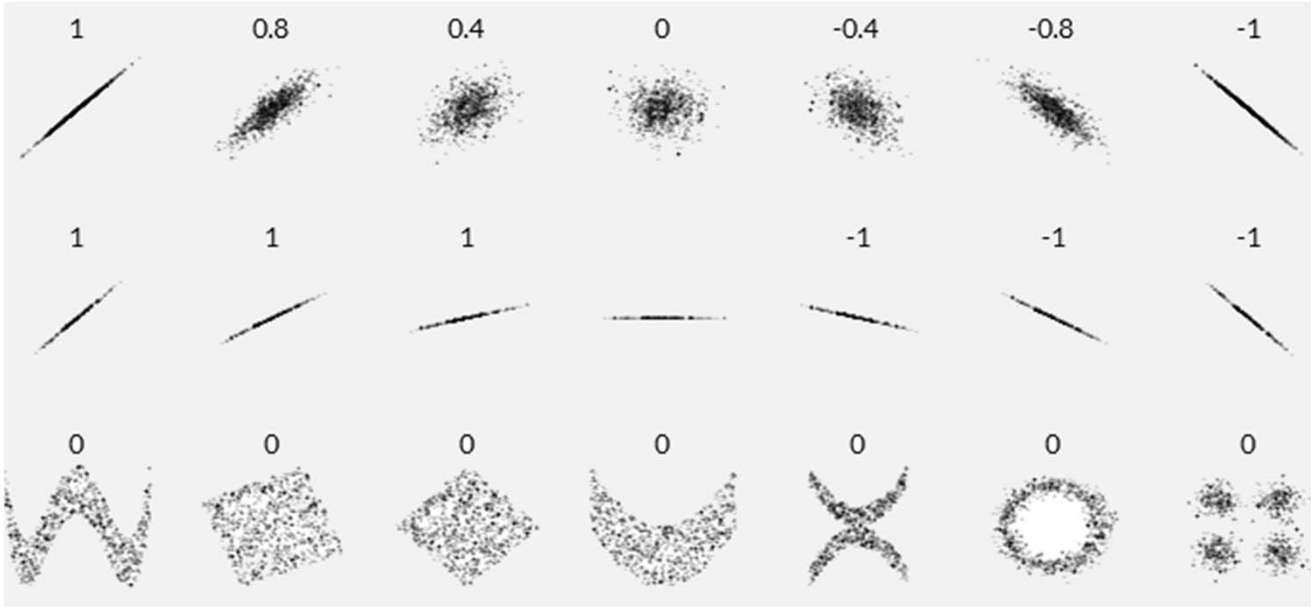
 # Example: find the covariance and the correlation coefficient from the given PDF.

- $f(X, Y) = 6xy(2 - x - y)$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$

Expectation, variance, covariance, correlation

(6) Correlation coefficient

Comparing correlation



Common distributions

(1) Normal distribution

A continuous random variable X is normally distributed with mean μ and variance σ^2, denoted as $X \sim N(\mu, \sigma^2)$, if the PDF is

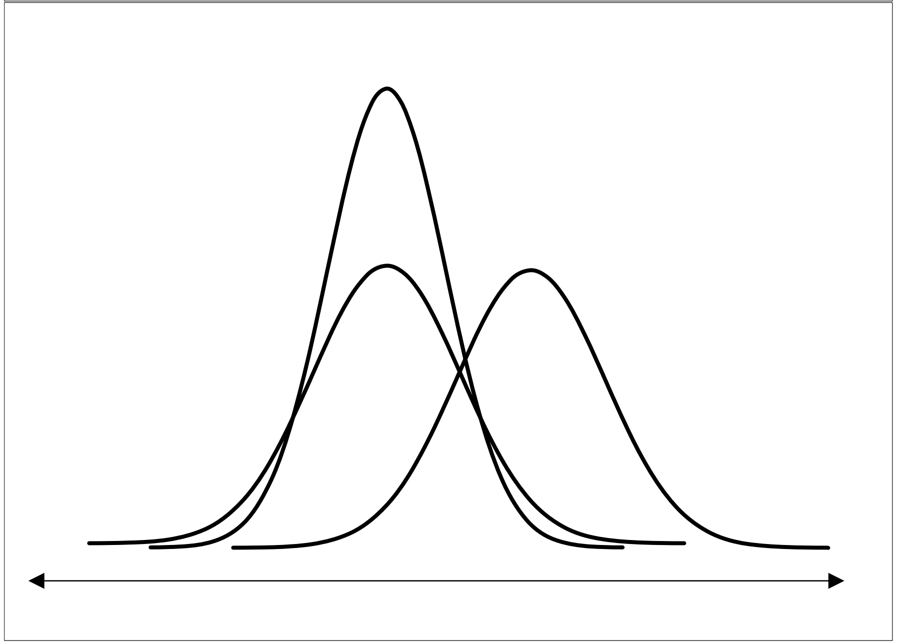
- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

If the mean or variance changes, the position and shape of the distribution also shift.

We can convert any X that is normally distributed into a **standard normal distribution**, defined as Z, by weighting as follows.

- $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

Normally distributed X



Common distributions

(2) Chi-square distribution

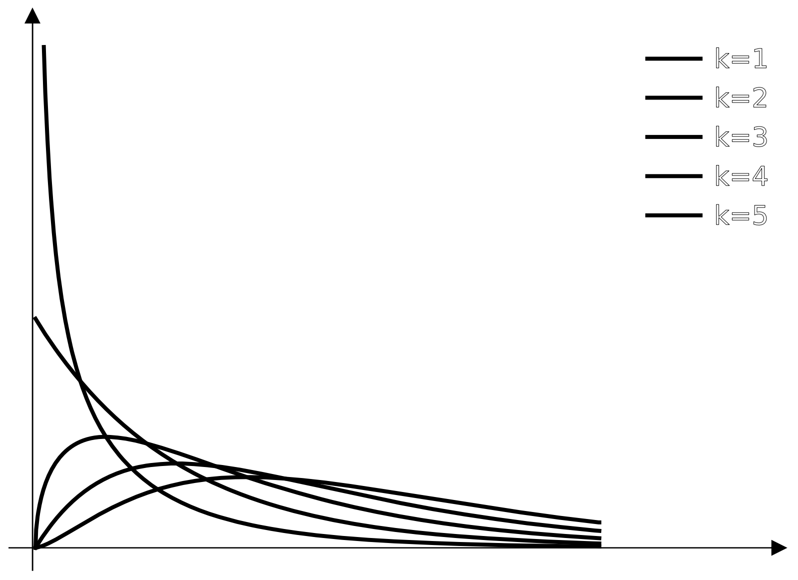
Chi-squared with k degrees of freedom (d.f.) is the distribution of a sum of the squares of k independent standard normal random variables.

- $\chi_k^2 = \sum_{i=1}^{k} Z_i^2$

Properties of χ_k^2

- Chi-square is skewed depending on d.f. As the d.f. increases it becomes more and more symmetrical.
- The mean is k and the variance is $2k$.
- If two chi-square variable is independent, the sum of them has the d.f. of $k_1 + k_2$.

Chi-square distribution with different d.f.



Common distributions

(3) Student's t-distribution

Let Z and χ_k^2 be random variables distributed as standard normal variable and chi-square respectively and they are independent, the t-distribution with k degrees of freedom can be represented as

- $t = \frac{Z\sqrt{k}}{\chi_k^2} \sim t_\nu$

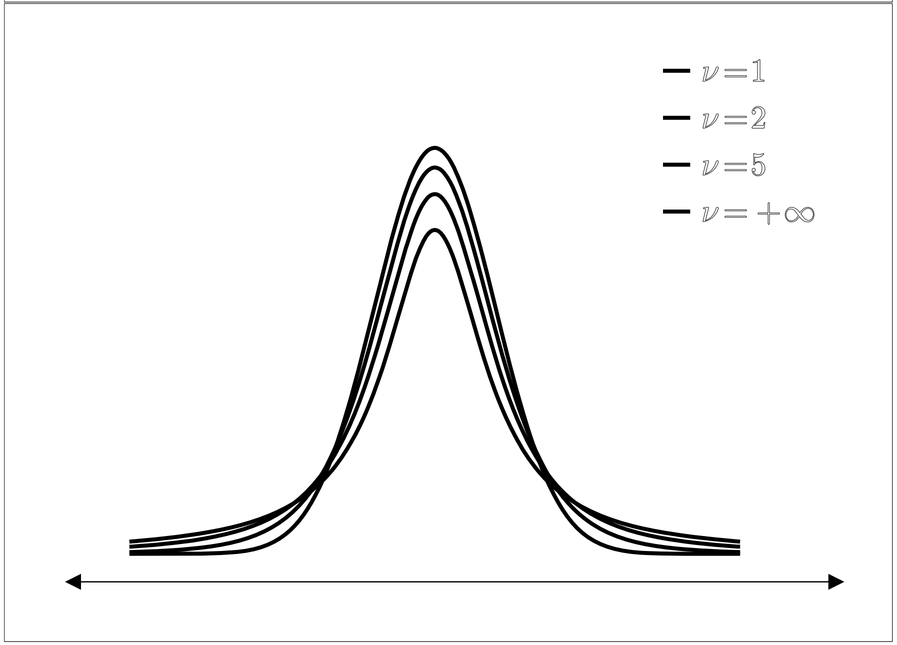
Properties of t

- The t-distribution is symmetrical but flatter compared to the normal distribution.

- As the d.f. increases, t-distribution is converted to the normal distribution.

- The mean is 0 and the variance is $\frac{k}{k-2}$.

t-distribution with different d.f.



Common distributions

(4) F-distribution

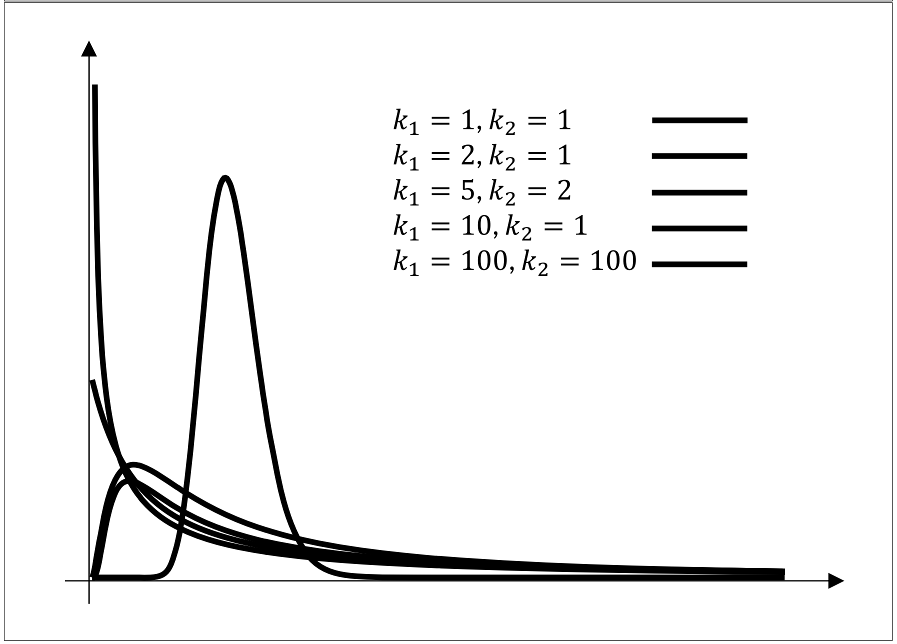
Let χ_1^2 and χ_2^2 be random variables distributed as chi-square and they are independent with the d.f. of k_1 and k_2, the F-distribution can be represented as

- $F = \frac{\chi_1^2/k_1}{\chi_2^2/k_2} \sim F(k_1, k_2)$

Properties of F

- The F-distribution is skewed to the right but if k_1 and k_2 becomes larger, the F-distribution becomes normal distribution.
- If k_2 is large, then we have $k_1 F \sim \chi_1^2$
- The square of t-distributed random variable with k degrees of freedom is $t_v^2 = F_{1,v}$
- The mean is $\frac{k_2}{k_2-2}$ and the variance is $\frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$.

F-distribution with different d.f.



Point estimation

An estimator, denoted by $\hat{\theta}$, is a statistics of a sample derived from a group of population. A decent estimator should have 2 properties.

- **Unbiased** – An estimator $\hat{\theta}$ is said to be an unbiased estimator of θ when

$$E(\hat{\theta}) = \theta.$$

We can further define the bias of the estimator as

$$bias(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

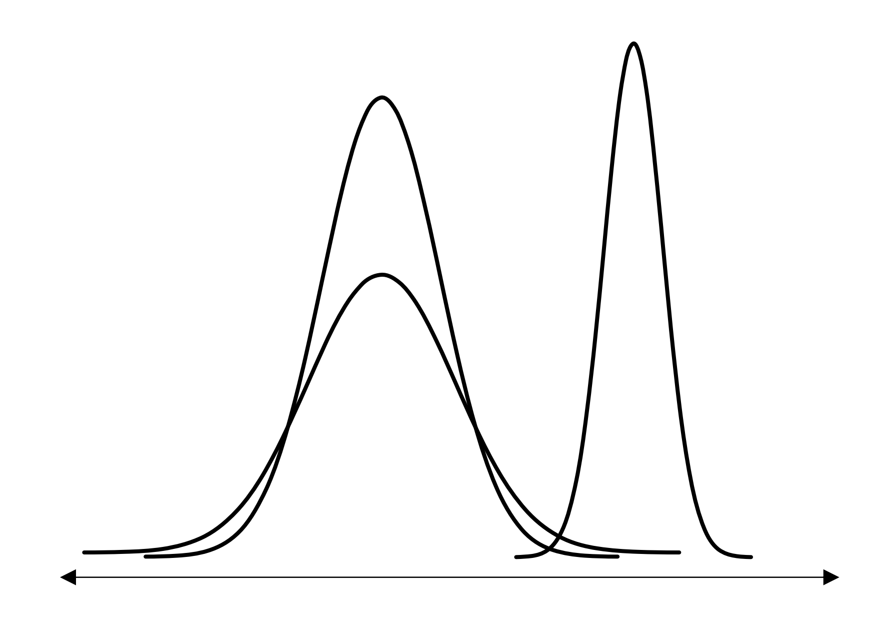
- **Efficient** – in other words, minimum variance. An estimator $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

If $\hat{\theta}$ is linear, unbiased, and has minimum variance in the class of all linear unbiased estimator of θ, then it is called the **Best Linear Unbiased Estimator (BLUE)**.

 # Example: find the $E(\bar{X})$ and $Var(\bar{X})$

Biasedness and efficiency of an estimator.



Large sample properties

Let $\hat{\theta}_n$ be an estimator of θ based on a random variable X, $\hat{\theta}$ is **consistent estimator** if

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

This is called **probability limit** (plim).

Now let X be independent, identically distributed (i.i.d.) random variables with mean μ then,

$$\text{plim}(\bar{X}) = \mu$$

This is also known as the **Law of Large Number** (LLN).

Consistent estimator