Probability and Statistics

Mhamed Ben Salah

The Graduate Institute of International and Development Studies, Geneva

Basic probability theory

Random variables

Common distributions

Random samples and limit theorems

Hypothesis testing

Basic probability theory

- The probability of an event is a positive number indicating how likely it is to happen.
- An event E is a subset of the different possible outcomes S. The probability of an event is the sum of the probabilities of its constituents outcomes.
- If A and B are events, then $A \cup B$ is the set of outcomes in at least one of A **or** B (at least one of A or B occurs). And $A \cap B$ is the set of outcomes that belong to A **and** B (both A and B occurs).

Basic probability theory

Some properties

- $0 \leq P(E) \leq 1$
- $P(S) = 1$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(\bar{A}) = 1 - P(A)$

Conditional probability $P(A|B)$ is the probability assigned to A knowing that B occurs. We have

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Basic probability theory

Baye's theorem Using the previous results we can show that

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$$

Independence The events A and B are said to be independent if

$$P(A \cap B) = P(A) \times P(B)$$

Example 1

Suppose a women in every 200 has a certain disease. A test for the disease shows positive for 99% of sufferers and 5% of non-sufferers. If a randomly chosen woman is tested for the disease and shows positive, what is the probability that she has the disease ?

Basic probability theory

An example of a probability law is the **Bernoulli trials**. The principle is to be able to know the probability of having a certain number k of success if you repeat an experiment n times. Let an event A be the fact of succeeding k time over n to the same experiment with a probability of success p at each try. The probability of A will be

$$P(A) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

N.B $\binom{n}{k}$ corresponds to the number of ways of choosing k elements from n, without regard to the order in which they are chosen

Basic probability theory

Example 2

Suppose you face a multiple choice format for your maths bootcamp final examination, with 30 questions and 4 possibilities for each question (with only one right answer). What is the probability of you obtaining the passing grade of 4 over 6 if you know absolutely nothing ?

Random variables

A **random variable** X is a real-valued function defined on a sample space S. In other words, it is a variable that can "randomly" take different values from a set of numbers (discrete or continuous).

The **cumulative distribution** (c.d.f.) F_X of a random variable X, is defined for each real number a by

$$F_X(a) = P(X \leq a)$$

Example 3

A fair coin is tossed until Head appears, let Y be the total number of tosses. Compute $F_Y(3)$.

Random variables

A **discrete random variable** is an r.v. whose range is either finite or countable. On the other hand a **continuous random variable** can take an infinite number of value (when we talk about time for instance). Because there are an infinite number of possibilities, the probability of a continuous variable X being exactly equal to any number x is null, $P(X = a) = 0$.

Let X be a continuous r.v. whose range is contained in $[a, b]$. The **density function** of x denoted f_X is defined by

$$F_X(x_0) = \int_a^{x_0} f_X(x)dx \quad \text{if } a \leq x_0 \leq b$$

And we have of course

$$\int_a^b f_X(x)dx = 1$$

Notes on Integration

Some rules on indefinte integrals

- $\int x^\alpha dx = \frac{1}{\alpha+1}x^{\alpha+1} + C \quad$ if $\alpha \neq -1$.
- $\int \frac{1}{x} dx = \ln x + C \quad$ if $x > 0$.
- $\int e^x dx = e^x + C$
- $\int (f(x) + g(x))dx = \int f(x)dx + \int g(x)dx$.
- If k is a constant, $\int kf(x) = k \int f(x)dx$.
- $\int (x-a)^\alpha dx = \frac{1}{\alpha+1}(x-a)^{\alpha+1} + C$
- $\frac{1}{x-a}dx = ln(x-a) + C$
- $\int e^{ax} dx = \frac{e^{ax}}{a} + C$

Notes on Integration

For the case of a **definite integrals** of the form

$$\int_a^b f(x)dx$$

We need to

- Calculate the indefinite integral $\int f(x)dx = F(x) + C$
- Calculate the definite integral (C cancels out)

$$[F(x)]_a^b = F(b) - F(a)$$

Random variables

Example 4

The density function of a continuous random variable X is given by

$$f_x(x) = \begin{cases} \alpha x(2-x) & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find α and $P(0.5 < X < 1.5)$.

Random variables

Expected value Let X be a **discrete** r.v. taking the values $a_1, a_2, \ldots, a_n$ with probabilities $p_1, p_2, \ldots, p_n$. The **expected value** of X, denoted by $E(X)$ or μ is defined as follows

$$\mu = E(X) = p_1 a_1 + p_2 a_2 + \ldots + p_n a_n$$

It can also be called the **expectation** or the **mean** of the variable.

If X is a **continuous** r.v. then

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx$$

Random variables

Variance The variance is a measure of dispersion of a random variable. It is the mean of the squared deviation of each variable from its the expected value. The variance is often denoted by σ^2 and can be expressed as follows

$$\sigma^2 = \text{var } X = E((X - EX)^2)$$

And this can be rewrite as

$$\begin{aligned}\sigma^2 &= E(X^2) - (EX)^2 \\ &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \mu^2\end{aligned}$$

The variance have the disadvantage of not being measured in the same units as X (it is the squared quantity of X). This is why we use another common measure of dispersion called the **standard deviation** denoted by σ, simply defined as $\sigma = \sqrt{\text{var } X}$

Random variables

Covariance and correlation The **covariance** of two random variables X and Y is

$$\begin{aligned}\operatorname{cov}(X,Y) &= E([X-EX][Y-EY]) \\ &= E(XY)-(EX)(EY)\end{aligned}$$

Notice that $\operatorname{cov}(X,Y) = \operatorname{var} X$ if $X = Y$. Moreover $\operatorname{cov}(X,Y) = 0$ if X and Y are independent.

The **correlation coefficient** between X and Y is

$$\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

with σ_X and σ_Y the standard deviations of X and Y.

Random variables

Some important properties of the **expected value**

- $E(a + bX) = E(a) + bE(X) = a + bE(X)$
- $E(X + Y) = E(X) + E(Y)$
- $E(X - Y) = E(X) - E(Y)$

and the **variance**

- $\text{var}(a + BX) = \text{var}(a) + \text{var}(bX) = b^2\text{var}(X)$
- $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$
- If X and Y are **independent** then
    - $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$
    - $\text{var}(X - Y) = \text{var}(X) \boldsymbol{+} \text{var}(Y)$

Random variables

Example 5

Let X be a r.v. with a density function

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

Find the expected value of X.

Random variables

Method of moments Let x be a r.v., for each integer k, $E(X^k)$ is called the kth **raw moment** of X and $E((X - EX)^k)$ is called the kth **central moment** of X.
The first raw moment is the **mean** and the second central moment is the **variance**.

When calculating moments of a r.v. X, we use the **moment generating function** of X, defined as

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= \int_{-\infty}^{+\infty} e^{tX} f_X(x) dx \end{aligned}$$

Random variables

The raw moment theorem   The reason why the moment generating function is useful is the raw moment theorem which states that

$$E(X^k) = \left[\frac{d^k}{dt^k} M_X(t)\right]_{t=0}$$

Example 6

Let X be a r.v. with a density function

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

Find the variance of X using the method of moments. (*Hint : you need to make the variable transformation u=x-t)*

Common distributions

Uniform distribution The probability remains the same ('uniform') on the entire interval. The uniform distribution associates the same probability to intervals of the same length. We have

$$f_x(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

For instance the density of a r.v. X uniformly distributed between $[0, 2]$ will be $f_X(x) = \frac{1}{2}$ on the same interval.

Normal distribution

One of the most common continuous r.v. is the **Gaussian** random variable, also called **standard normal variate**. Its density function, usually denoted by ϕ is

$$\phi(x) = f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$$

The corresponding c.d.f is $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$.

The curve $y = \phi(x)$ is symmetrical about the vertical axis and the area under the curve is 1. Therefore we have $\phi(0) = \frac{1}{2}$.

As we proved in examples 5 and 6, $E(X) = 0$ and $\text{var}(X) = 1$. Therefore the fact that X follows a standard normal distribution can be written $X \sim \mathcal{N}(0, 1)$

Normal distribution

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

FIGURE 1 – Table of cumulative normal distribution $\Phi(x)$

Normal distribution

Standardisation utility of the normal distribution lies in the fact that any normal variable, with any expected value μ and any variance σ^2 can be **standardised** to a standard normal variate.

Consider a normal r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$, we can always transform X in a r.v Z such that $Z = \frac{X-\mu}{\sigma}$. The expected value and the variance of Z will be

- $$E(Z) = E(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma}(E(X) - \mu) = \frac{1}{\sigma}(\mu - \mu) = 0$$
- $$\text{var}(Z) = \text{var}(\frac{X-\mu}{\sigma}) = \frac{1}{\sigma^2}(\text{var}(X) + \text{var}(\mu)) = \frac{1}{\sigma^2} \cdot \sigma^2 = 1$$

Therefore we have $Z \sim \mathcal{N}(0, 1)$

Normal distribution

Example 7

Consider the random variable $X \sim \mathcal{N}(3, 4)$. Find

- $P(X \geq 6)$
- $P(X \leq 1.5)$
- $P(1 \leq X \leq 4)$

Random samples and limit theorems

- The r.v. $X_1, X_2, \ldots, X_n$ are said to be **identically distributed** if they have the same c.d.f. If these r.v. are identically distributed and independent (*iid*) they are said to form a **random sample**.
- For example, suppose we sample at random 200 taxi-drivers form the whole population of all Geneva taxi-drivers. Let X_i be the earnings of th ith driver sampled. Assuming that the earning of a taxi driver is independent from a randomly chosen any other taxi driver, then $X_1, X_2, \ldots, X_{200}$ form a random sample.

Random samples and limit theorems

- Because X_i has the same c.d.f. for all i, EX_i is the same for all i. This expected value is the mean of the entire population's earnings, we call it the **population mean** and denote it by μ. Same logic for the **population variance** σ^2. The **sample mean** $\bar{X}$ is defined by

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$$

It is important to notice that the sample mean $\bar{X}$ is a random variable, whereas μ is not.

$$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\mu = \mu$$

$$\text{var}(\bar{X}) = \text{var}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \text{var}(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Random samples and limit theorems

Before stating the law of large number, we need to define another important theorem, known as the **Chebyshev Inequality**.

Chebyshev Inequality   Let X be a discrete random variable with expected value $\mu = E(X)$, variance $V(X)$, and let $\epsilon > 0$ be any positive real number. Then

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

Random samples and limit theorems

Law of large numbers For any positive number γ, however small, the probability that $\bar{X}_n$ differs from μ by less than γ approaches 1 as $n \to \infty$

$$\lim_{n\to\infty} \quad P(\mu - \gamma < \bar{X}_n < \mu + \gamma) = 1 \quad \text{for all } \gamma > 0$$

It means that if n is large enough, the r.v. $\bar{X}_n$ is probably close to μ (and not only its expected value).

Proof Since $X_1, X_2, \ldots, X_n$ are independent random variables identically distributed with $E(X) = \mu$ and $V(X) = \sigma^2$ we know that the variance of $S_n = \sum_{i=1}^n X_i$ is

$$V(S_n) = n\sigma^2 \qquad \text{and} \qquad V(\frac{S_n}{n}) = V(\bar{X}_n) = \frac{\sigma^2}{n}$$

Random samples and limit theorems

We also know that

$$E(\bar{X}_n) = \mu$$

By Chebyshev's Inequality we have, for any $\gamma > 0$

$$P(|\bar{X}_n - \mu| \geq \gamma) \leq \frac{\sigma^2}{n\gamma^2}$$

Therefore, as $n \to \infty$

$$P(|\bar{X}_n - \mu| \geq \gamma) \to 0$$

Which means that, as $n \to \infty$

$$\begin{aligned} & P(|\bar{X}_n - \mu| < \gamma) \to 1 \\ \Rightarrow \quad & P(-\gamma < \bar{X}_n - \mu < \gamma) \to 1 \\ \Rightarrow \quad & P(\mu - \gamma < \bar{X}_n < \mu + \gamma) \to 1 \end{aligned}$$

N.B. This law is often called the **weak law of large numbers**, the **strong law of large numbers**, which is harder to prove, has a stricter definition of 'probably close'.

Example 8

A fair coin is tossed 100 times. The expected number of heads is 50, and the standard deviation $\sigma = 5$. What does Chebyshev's Inequality tell you about the probability that the number of heads that turn up deviates from the expected number by three or more standard deviations ?

Random samples and limit theorems

The central limit theorem Let $\bar{X}_n$ be the sample mean of a random sample. For each n, let

$$Z_n = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{var}(\bar{X}_n)}}$$

Then $E(Z_n) = 0$ and $\text{var}(Z_n) = 1$. The **central limit theorem** states that, as $n \to \infty$, the c.d.f. of Z_n approaches that of a standard normal variate. In other words, for large n, the sample mean $\bar{X}_n$ is approximately normally distributed and therefore can be standardised into a $\mathcal{N}(0, 1)$.

N.B. The approximation usually works well if $n \geq 50$.

Example 9

Let W be the sum of 50 independent random variables, each uniformly distributed over the interval $[0, 2]$. Find $P(48 < W < 52)$

Hypothesis testing

- An hypothesis is defined as an assumption about the **population**. We often have hypotheses about the population mean μ. The intuition is that we use the information we have about our sample, mostly the sample mean $\bar{X}$, in order to test our hypothesis about the population.
- The hypothesis to be tested is called the **null hypothesis** H_0. The hypothesis against which the null is tested is called the **alternative hypothesis** H_A.

$$\begin{aligned} H_0 : &\quad \mu = \mu_0 \\ H_A : &\quad \mu \neq \mu_0 \end{aligned}$$

- Two assumptions have to be made :
  - $\bar{X}$ is normally distributed (LLN and the CLT).
  - The population variance σ^2 is known.

Hypothesis testing

- Therefore we assume that our random variable $\bar{X}$ is normally distributed with mean μ_0 and variance $\frac{\sigma^2}{n}$. If the true mean is μ_0, the values of $\bar{X}$ that are "close" to μ_0 will occur with greater probability than the values "farther away" from μ_0.
- We would **reject the null hypothesis** if the value of $\bar{X}$ turns out to be *so low* or *so high* compared to μ_0 that its occurrence by chance would be very "unlikely".
- We call the **significance level** of a test, denoted by α, the probability for the random variable to be equal to a given value that we consider to be too low to consider that the random variable is equal to its mean. We usually test for significance levels of 1%, 5% or 10%.

Hypothesis testing

We saw earlier that it is easier to work with standard normal variables. Therefore we start by standardising our r.v. $\bar{X}$.

$$z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}$$

This will be our **test statistic**.

Then, given the significance level, we look for the **critical values**, the values over which our test statistic is to high and therefore we reject the null hypothesis. The critical value is the value t_α such that

$$P(Z > t_\alpha) = \alpha$$

For $\alpha = 0.05,\ t_\alpha = 1.64$; for $\alpha = 0.025,\ t_\alpha = 1.96$.

Hypothesis testing

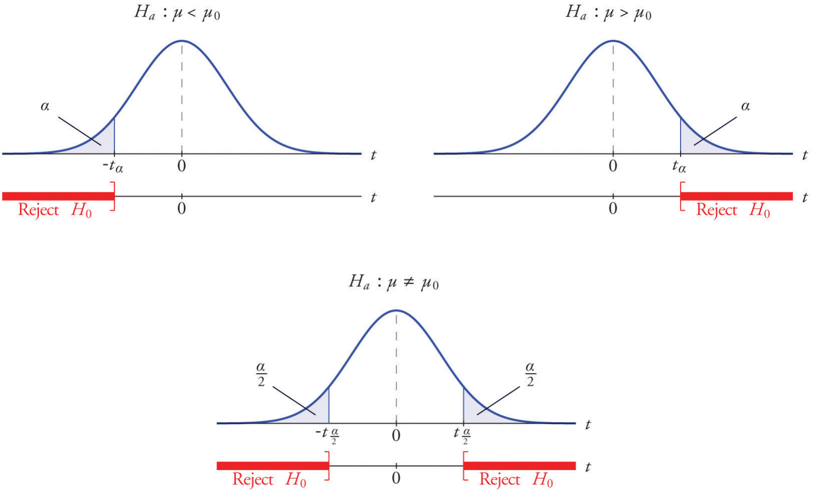

FIGURE 2 – Distribution of the standardised $\bar{X}$ and significance levels

Hypothesis testing

If the test statistic z is higher than the critical value ($z > t_\alpha$), then we can **reject the null hypothesis** and say for instance that μ is not equal to μ_0. However if $z < t_\alpha$ does not mean that $\mu = \mu_0$, we can only say that we **fail to reject** the fact that they are equal.

In hypothesis testing we can do two kind of mistakes :

- **Type I Error** is rejecting H_0 when it is true. The probability of doing this mistake is α. The smallest α is, the lower the probability of Type I error will be and the more "powerful" will be our test.
- **Type II error** is failing to reject H_0 when it is false. The power of a test is one minus the probability of a Type II error.

Hypothesis testing

A small recap :

- We first set our hypothesis about the population

$$\begin{aligned} H_0: &\quad \mu = \mu_0 \\ H_A: &\quad \mu \neq \mu_0 \end{aligned}$$

- We assume (often using LLN and CLT) that X is normally distributed and that we know the population variance σ^2. Hence we are assuming that our sample mean $\bar{X} \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{n})$
- We standardise our sample mean in order to get our test statistic z

$$z = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}$$

Hypothesis testing

- We look for the critical value t_α depending on the significance level of the test and the alternative hypothesis.
- We compare our test statistic to the critical value

$$\begin{cases} |z| > |t_\alpha| & \text{We reject } H_0 \\ |z| < |t_\alpha| & \text{We fail to reject } H_0 \end{cases}$$

Hypothesis testing

Example 10

Psychological studies indicate that in the population at large intelligence, IQ is normally distributed with mean 100 and a standard deviation of 16. Suppose we want to test whether the left handed are characterized by a different mean.
We have at our disposal a sample of 400 left handed for which the mean IQ is 102. Assuming that the standard deviation among the left-handed population is also equal to 16, test with a significance level of 5% whether the IQ of the left-handed is different from the rest of the population.