BATNA 2 University of Algeria.
Mathematics and Computer Science Faculty
Common Core in Mathematics and Computer Science Department
Semester-3. L2 SCMI.

Probabilities and Statistics II

Course handout
For the use of students in second year.

Dr. Hiba Aiachi

Academic year (2023/2024)

Contents

1 Expectation **7**
- 1.1 The Discrete Case . . . 7
- 1.2 The Absolutely Continuous Case . . . 9
- 1.3 Properties . . . 9
- 1.4 Variance, Covariance, and Correlation . . . 10
- 1.5 Conditional Expectation . . . 16
  - 1.5.1 Discrete case . . . 16
  - 1.5.2 Absolutely Continuous Case . . . 19

2 Sampling Distributions and Limits **21**
- 2.1 Sampling Distributions . . . 21
- 2.2 Convergence in Probability . . . 22
  - 2.2.1 The Weak Law of Large Numbers . . . 24
- 2.3 Convergence Almost Surely . . . 25
  - 2.3.1 The Strong Law of Large Numbers . . . 26
- 2.4 Convergence in Distribution . . . 27
  - 2.4.1 The Central Limit Theorem . . . 29
  - 2.4.2 Cumulative Distribution Function of the Standard Normal Distribution . . . 31
- 2.5 Monte Carlo Approximations . . . 34

3 Statistical Inference **35**
- 3.1 Statistical model . . . 35
- 3.2 Point estimation . . . 36
  - 3.2.1 Method of moments . . . 37
  - 3.2.2 Maximum of likelihood estimation . . . 41
- 3.3 Confidence interval estimation . . . 43
  - 3.3.1 Classic examples of interval estimation . . . 43

4 Statistical tests **50**
4.1 Homogeneity tests 50
4.1.1 Comparison of two means (Test of Student) 50
4.1.2 Comparison of two variances (Test of Fisher) 55

Review

Definition 0.0.1. *(Random variable)*
A random variable (r. v) is a function from the sample space Ω *to the set of all real numbers* $\mathbb{R}$. *i.e.*

$$\begin{aligned} X : \Omega &\longmapsto \mathbb{R} \\ \omega &\longmapsto X(\omega) \end{aligned}$$

Exemple 0.0.1. *If the sample space corresponds to flipping two different coins, then we could let* X *be the total number of heads showing, so;*

$$X(\Omega) = \{0, 1, 2\}$$

ω	TT	TH	HT	HH
$X(\omega)$	*0*	*1*	*1*	*2*

Definition 0.0.2. *(Cumulative Distribution Function)*
Given a random variable X*, its cumulative distribution function (or cdf for short) is the function* F_X *defined by*

$$\begin{aligned} F_X : \mathbb{R} &\longmapsto [0, 1] \\ t &\longmapsto F_X(t) = P[X \leq t] \end{aligned}$$

Where there is no confusion, we sometimes write $F(t)$ *for* $F_X(t)$.

Properties of cumulative distribution Function

1. $0 \leq F_X(t) \leq 1$. for all t
2. $F(x) \leq F(y)$ whenever $x \leq y$ (i.e., F_X is increasing),
3. $\underset{t \to -\infty}{lim} F_X(t) = 0$, $\underset{t \to +\infty}{lim} F_X(t) = 1$.

4. If $a \leq b$ $P[a \leq X \leq b] = F_X(b) - F_X(a)$.

Definition 0.0.3. *(Discrete random variables) A random variable X is discrete if $X(\Omega)$ is a finite or countable sequence of distinct real numbers, and a corresponding sequence.*

Definition 0.0.4. *For a discrete random variable X, its probability function is the function*

$$P : X(\Omega) \longmapsto [0,1]$$
$$x_i \longmapsto P[X = x_i]$$

where $X(\Omega) = \{x_1, x_2, ..., x_n\}$, such that $P(X = x_i) = pi$ for all i, we have $\sum_i p_i = 1$.

Hence,

1. for all x $:F_X(x) = \sum_{x_i \leq x} P[X = x_i]$.
2. The distribution function of X is a staircase function, it jumps to non-zero probability points.

Important Discrete Distributions

Exemple 0.0.2. *(Bernoulli Distribution) Consider flipping a coin that has probability θ of coming up heads and probability $1 - \theta$ of coming up tails, where $0 \leq \theta \leq 1$.*
Let $P_X(1) = P(X = 1) = \theta$, while $P_X(0) = P(X = 0) = 1 - \theta$. The random variable X is said to have the Bernoulli distribution, we write this as $X \rightsquigarrow Bernoulli(\theta)$.

Exemple 0.0.3. *(The Binomial Distribution) the Binomial distribution is the sum of n independent Bernoulli distributions with parameter θ, such that we consider flipping n coins with independent probability θ of coming up heads and probability $1 - \theta$ of coming up tails, (Again $0 \leq \theta \leq 1$). Let X be be the total number of heads showing. we obtain then*

$$P[X = k] = C_n^k \theta^k (1 - \theta)^{n-k} \text{ where } C_n^k = \frac{n!}{k!(n-k)!}$$

The random variable X is said to have the Binomial distribution we write this as $X \rightsquigarrow Binomial(n, \theta)$

Exemple 0.0.4. *(The Geometric Distribution)*
Consider repeatedly flipping a coin that has probability θ of coming up heads and probability $1 - \theta$ of coming up tails, where again $0 \leq \theta \leq 1$. Let X be the number of tails that appear before the first head. Then for $k \geq 0, X = k$ if and only if the coin shows

exactly k tails followed by a head. The probability of this is equal to $(1-\theta)^k\theta$. (In particular, the probability of getting an infinite number of tails before the first head is equal to $(1-\theta)^\infty\theta = 0$, so X is never equal to infinity.) Hence, $P_X(k) = (1-\theta)^k\theta$,for $k = 0, 1, 2, 3, ...$ The random variable X is said to have the Geometric distribution, we write this as $X \rightsquigarrow Geometric(\theta)$.

Exemple 0.0.5. *(Poisson Distribution)*
Suppose that X has the Binomial distribution (n, θ). Then for $0 \le x \le n$, If we set $\theta = \frac{\lambda}{n}$ for some value $\lambda > 0$ and Let us now consider that $n \to \infty$ while keeping x fixed at some non-negative integer.

$$\lim_{n\to\infty} P[X = k] = \frac{1}{k!}\exp(-\lambda)\lambda^k$$

We can phrase this result as follows. If we flip a very large number of coins n, and each coin has a very small probability $\theta = \frac{\lambda}{n}$ of heads coming up, then the probability that the total number of heads will be k is approximately given by $\exp(-\lambda)\lambda^k/k!$. The random variable X is said to have the Poisson distribution, we write this as $X \rightsquigarrow Poisson(\lambda)$

Exemple 0.0.6. *(The Hypergeometric Distribution)*
the hypergeometric distribution will apply to any context wherein we are sampling without replacement from a finite set of N elements and where each element of the set either has a characteristic or does not. For example, if we randomly select people to participate in an opinion poll so that each set of n individuals in a population of N has the same probability of being selected, then the number of people who respond yes to a particular question is distributed Hypergeometric(N, M, n) where M is the number of people in the entire population who would respond yes.
Hence,

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$

Definition 0.0.5. *(Continuous random variables)*
A random variable X is continuous if $X(\Omega)$ is an interval (part of $\mathbb{R}$) so $P(X = x) = 0$

Definition 0.0.6. *We call a function $f : \mathbb{R} \to \mathbb{R}$ a density function if*
For all $x \in \mathbb{R}$: $f(x) \ge 0$
$\int_{-\infty}^{+\infty} f(x)dx = 1$

Definition 0.0.7. *A random variable X is absolutely continuous if there is a density function f , such that*

$$P(a \le X \le b) = \int_a^b f(x)dx$$

whenever $a \leq b$
An absolutely continuous random variable is a continuous variable.

Definition 0.0.8. *The cumulative distribution function of a continuous random variable is defined by*

$$\forall x \in \mathbb{R} : F_X(x) = \int_{-\infty}^{x} f(t)dt$$

Important Absolutely Continuous Distributions

Exemple 0.0.7. *(The Uniform Distribution)*
Let L and R be any two real numbers with $L, R \in \mathbb{R}$. Consider a random variable X such that

$$f(x) = \begin{cases} \frac{1}{R-L} & L \leq x \leq R \\ 0 & otherwise \end{cases}$$

The random variable X is said to have the Uniform$[L, R]$ distribution, we write this as $X \rightsquigarrow Uniform[L, R]$.

Exemple 0.0.8. *(Gauss distribution (Normal distribution))*
Let X be a random variable having the density function given by

$$\forall x \in \mathbb{R}, \mu \in \mathbb{R}; \sigma \in \mathbb{R}^+, f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2)$$

The random variable X is said to have the normal distribution noted by $N(\mu, \sigma^2)$.
If $\mu = 0$ and $\sigma = 1$, then this corresponds with the standard normal distribution, we write this as $X \rightsquigarrow N(0, 1)$. then the density function of X is

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

Exemple 0.0.9. *(Distributions derived from the normal distribution)*
• *Let these random variables $X_1, \ldots, X_n$ be independent and distributed to a standard normal distribution, for all i we have, $X_i \rightsquigarrow N(0, 1)$*
If $S = X_1^2 + X_2^2 + \cdots + X_n^2$, then S have the Chi squared distribution which has n degrees of freedom.
We note $S \rightsquigarrow \chi^2(n)$.
• *Let the independent random variables U, V, where $U \rightsquigarrow \mathcal{N}(0; 1)$ and $V \rightsquigarrow \chi^2_{n.d.d.l}$,*
If $T_n = \frac{U}{\sqrt{\frac{V}{n}}}$ then T_n follows the Student distribution with n degrees of freedom.
Notation: $T_n \rightsquigarrow T_n$

Chapter 1

Expectation

the expected value of a random variable is the average value that the random variable takes on. For example, if half the time $X = 0$, and the other half of the time $X = 10$, then the average value of X is 5. We shall write this as $E(X) = 5$. Similarly, if one third of the time $Y = 6$ while two thirds of the time $Y = 15$, then $E(Y) = 12$.
To understand expected value more precisely, we consider discrete and absolutely continuous random variables separately.

SECTION 1.1

The Discrete Case

Definition 1.1.1. *Consider a random variable X with a finite list $x_1, ..., x_k$ of possible outcomes, each of which (respectively) has probability $P_1, ..., P_k$ $(i = 1, ..., kP_i = P(X = x_i))$ of occurring. The expectation of X is defined as*

$$\mathrm{E}(X) = x_1P_1 + x_2P_2 + \cdots + x_kP_k = \sum_{i=1}^{k} x_iP_i$$

Exemple 1.1.1. *Let X represent the outcome of a roll of a fair six-sided die. The possible values for X are 1, 2, 3, 4, 5, and 6, all of which are equally likely with a probability of 1/6. The expectation of X is*

$$\mathrm{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

Exemple 1.1.2. *Suppose that $P(Y = -3) = 0.2$, and $P(Y = -11) = 0.7$, and $P(Y = 31) = 0.1$. Then*

$$\mathrm{E}(Y) = (-3)0.2 + (-11)0.7 + (31)0.1 = -0.6 - 7.7 + 3.1 = -5.2$$

In this case, the expected value of Y is negative

We now consider some of the common discrete distributions introduced above

Exemple 1.1.3. *(Bernoulli distribution)*
If $X \rightsquigarrow Bernoulli(\theta)$, then $P(X = 1) = \theta$ and $P(X = 0) = 1 - \theta$, so

$$E(X) = (1)\theta + 0(1 - \theta) = \theta$$

Exemple 1.1.4. *(Binomial distribution)*
If $Y \rightsquigarrow Binomial(n, \theta)$, then $P(Y = k) == C_n^k \times \theta^k \times (1-\theta)^{n-k}$, for $k = 1, ..., n$. Hence

$$\begin{aligned} E(Y) &= \sum_{k=1}^{n} kP(Y = k) = \sum_{k=1}^{n} kC_n^k \theta^k (1-\theta)^{n-k} \\ &= \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \theta^k (1-\theta)^{n-k} \\ &= \sum_{k=1}^{n} \frac{n(n-1)!}{(k-1)!(n-k)!} \theta^k (1-\theta)^{n-k} = \sum_{k=1}^{n} nC_{n-1}^{k-1} \theta^k (1-\theta)^{n-k} \end{aligned}$$

Now, the binomial theorem says that for any a and b and any positive integer m,

$$(a+b)^m = \sum_{j=0}^{m} C_m^j a^j b^{m-j}$$

Using this, and setting $j = k - 1$, we see that

$$\begin{aligned} E(Y) &= \sum_{k=1}^{n} nC_{n-1}^{k-1} \theta^k (1-\theta)^{n-k} = \sum_{j=0}^{n-1} nC_{n-1}^{j} \theta^{j+1} (1-\theta)^{n-1-j} \\ &= n\theta \sum_{j=0}^{n-1} C_{n-1}^{j} \theta^{j} (1-\theta)^{n-1-j} \\ &= n\theta \left[(\theta + 1 - \theta)^{n-1} \right] \end{aligned}$$

Hence $E(Y) = n\theta$.

SECTION 1.2

The Absolutely Continuous Case

Definition 1.2.1. *Let X be an absolutely continuous random variable, with density function $f(x)$. Then the expected value of X is given by*

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

From this definition, it is not too difficult to compute the expected values of many of the standard absolutely continuous distributions.

Exemple 1.2.1. *(The Uniform Distribution)*
Let $X \rightsquigarrow Uniform[L, R]$ so that the density of X is given by

$$f(x) = \begin{cases} \frac{1}{R-L} & L \leq x \leq R \\ 0 & otherwise \end{cases}$$

Hence,

$$\begin{aligned} E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_{L}^{R} x \frac{1}{R-L} dx &= \left. \frac{x^2}{2(R-L)} \right]_{x=L}^{x=R} \\ = \frac{R^2 - L^2}{2(R-L)} = \frac{(R-L)(R+L)}{2(R-L)} &= \frac{R=L}{2} \end{aligned}$$

Thus $E(X) = \dfrac{R=L}{2}$

SECTION 1.3

Properties

1. If $X = Y$, then $E(X) = E(Y)$ In other words, if X and Y are random variables that take different values with probability zero, then the expectation of X will equal the expectation of Y.

2. If $X = c$ for some real number c, then $E(X) = c$. In particular, for a random variable X with well-defined expectation, $E[E(X)] = E(X)$. A well defined expectation implies that there is one number, or rather, one constant that defines the expected value. Thus follows that the expectation of this constant is just the original expected value.

Theorem 1.3.1. *(Law of the unconscious statistician)*
Let X be a random variable, and let $g : \mathbb{R} \to \mathbb{R}$ be some function such that the expectation of the random variable $g(X)$ exists. Then

$$E[g(x)] = \sum_x g(x)P(X = x) \left(E[g(x)] = \int_{\mathbb{R}} g(x)f(x) \right)$$

Theorem 1.3.2. *(Linearity of expected values) Let X and Y be random variables, and let a and b be real numbers, and put $Z = aX + bY$. Then*

$$E(Z) = aE(X) + bE(Y)$$

.

Theorem 1.3.2 says, in particular, that $E(X + Y) = E(X) + E(Y)$, i.e., that expectation preserves sums. It is reasonable to ask whether the same property holds for products. That is, do we necessarily have $E(XY) = E(X)E(Y)$? In general, the answer is no, as the following example shows.
On the other hand, if X and Y are independent, then we do have $E(X)E(Y) = E(XY)$.

Theorem 1.3.3. *Let X and Y be independent random variables. Then $E(XY) = E(X)E(Y)$.*

Theorem 1.3.3 will be used often in subsequent chapters, as will the following important property.

Theorem 1.3.4. *(Monotonicity)*
Let X and Y be discrete random variables, and suppose that $X \leq Y$. (Remember that this means $X(s) \leq Y(s)$ for all $s \in \Omega$). Then $E(X) \leq E(Y)$.

SECTION 1.4

Variance, Covariance, and Correlation

Now that we understand expected value, we can use it to define various other quantities of interest. The numerical values of these quantities provide information about the distribution of random variables.
Given a random variable X, we know that the average value of X will be $E(X)$. However, this tells us nothing about how far X tends to be from $E(X)$. For that, we have the following definition.

Definition 1.4.1. *The variance of a random variable X is the quantity*

$$\sigma^2 = Var(X) = E\left((X - \mu_X)^2\right)$$

Where $\mu_X = E(X)$ is the mean of X

We note that it is also possible to write $Var(X) = E\left((X - E(X))^2\right)$, however the multiple uses of "E" may be confusing.

Remark 1.4.1. *We often use the variance in the form*

$$Var(X) = E(X^2) - \mu_X^2$$

(That is, variance is equal to the second moment minus the square of the first moment.)

Proof. Using the property 2and the theorem 1.3.2 we could clearly find that

$$\begin{aligned} E(X) =& E((X - \mu_X)^2) = E(X^2 - 2X\mu_X + \mu_X^2) \\ &= E(X^2) - 2\mu_X E(X) + \mu_X^2 = E(X^2) - 2\mu_X^2 + \mu_X^2 \\ &= E(X^2) - \mu_X^2 \end{aligned}$$

as claimed. □

Also, because $(X - \mu_X)^2$is always non-negative, its expectation is always defined, so the variance of X is always defined. Intuitively, the variance $Var(X)$ is a measure of how spread out the distribution of X is, or how random X is, or how much X varies, as the following example illustrates.

Exemple 1.4.1. *Let X and Y be two discrete random variables, with probability functions*

$$P(X = x) = \begin{cases} 1 & x = 10 \\ 0 & otherwise \end{cases}$$

and

$$P(Y = y) = \begin{cases} 1/2 & y = 5 \\ 1/2 & y = 15 \\ 0 & otherwise \end{cases}$$

respectively.
Then $E(X) = E(Y) = 10$. However,

$$Var(X) = E(X^2) - E(X)^2 = 100 - 100 = 0$$

while

$$Var(Y) = E(Y^2) - E(Y)^2 = 5^2(1/2) + 15^2(1/2) - 10^2 = 25$$

We can see that, while X and Y have the same expected value, the variance of Y is much greater than that of X. This corresponds to the fact that Y is more random than X, that is, it varies more than X does.

Exemple 1.4.2. *Let $Y \rightsquigarrow Bernoulli(\theta)$. Then $E(Y) = \theta$. Hence,*

$$Var(Y) = E(Y^2) - E(Y) = 1^2\theta + 0^2(1-\theta) - \theta^2 = \theta - \theta^2 = \theta(1-\theta)$$

The square in the definition 1.4.1 implies that the "scale" of $Var(X)$ is different from the scale of X. For example, if X were measuring a distance in meters (m), then $Var(X)$ would be measuring in meters squared (m^2). If we then switched from meters to millimeter, we would have to multiply X by 1000 but would have to multiply $Var(X)$ by about $(1000)^2$.
To correct for this "scale" problem, we can simply take the square root, as follows.

Definition 1.4.2. *The standard deviation of a random variable X is the quantity*

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{E(X^2) - E(X)^2}$$

It is reasonable to ask why, in 1.4.1, we need the square at all. Now, if we simply omitted the square and considered $E(X - \mu_X)$, we would always get zero (because $\mu_X = E(X)$), which is useless. On the other hand, we could instead use $E(|X - \mu_X|)$. This would, like 1.4.1, be a valid measure of the average distance of X from μ_X. Furthermore, it would not have the "scale problem" that $Var(X)$ does. However, we shall see that $Var(X)$ has many convenient properties. By contrast, $E(|X - \mu_X|)$ is very difficult to work with. Thus, it is purely for convenience that we define variance by $E((X - \mu_X)^2)$ instead of $E(|X - \mu_X|)$.
The variance has a great importance. Thus, we pause to present some important properties of it.

Theorem 1.4.1. *Let X be any random variable, with expected value $\mu_X = E(X)$, and variance $Var(X)$. Then the following hold true:*
(a)$Var(X) \geq 0$.
(b) If a and b are real numbers, $Var(aX + b) = a^2 Var(X)$.
(c) $Var(X) \leq E(X^2)$.

Proof. (a) This is immediate, because we always have $(\mu_X - X)^2 \geq 0$.

(b) We note that $\mu_{aX+b} = E(aX+b) = aE(X)+b = a\mu_X + b$, by linearity. Hence, again using linearity,

$$\begin{aligned} Var(aX+b) =& E((aX+b-\mu_{aX+b})^2) = E((aX+b-a\mu_X-b)^2) \\ &= E(a^2(X-\mu_X)^2) = a^2E((X-\mu_X)^2) \\ &= a^2Var(X) \end{aligned}$$

(d) This follows immediately from part (c) because we have $-\mu_X^2 \leq 0$. □

Corollary 1.4.1. *Let X be any random variable, with standard deviation $\sigma(X)$, and let a be any real number. Then $\sigma(aX) = |a|\sigma(X)$.*

Exemple 1.4.3. *Variance and Standard Deviation of the $N(\mu, \sigma^2)$ Distribution Suppose that $X \rightsquigarrow N(\mu, \sigma^2)$. Before we established that $E(X) = \mu$ Now we compute $Var(X)$. First consider $Z \rightsquigarrow N(0, 1)$ Then we have that*

$$Var(Z) = E(Z^2) = \int_{-\infty}^{+\infty} \frac{z^2}{\sqrt{2\pi}} \exp(-\frac{z^2}{2})dz$$

Then using integration by parts, we put $u = z$ and $v' = z\exp(-\frac{z^2}{2})$ so($u' = 1$ and $v = -\exp(-\frac{z^2}{2})$), we obtain

$$E(Z^2) = \frac{1}{\sqrt{2\pi}} z\exp(-\frac{z^2}{2})\Big]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2})dz = 1$$

and $\sigma(Z) = 1$
Now, for $\sigma > 0$ put $X = \mu Z + \sigma$ We then have $X \rightsquigarrow N(\mu, \sigma^2)$. From Theorem 1.4.1(b) we have that

$$Var(X) = \sigma^2 Var(Z) = \sigma^2$$

and $\sigma(X) = \sigma$ This establishes the variance of the $N(\mu, \sigma^2)$ distribution as σ^2 and the standard deviation as σ

So far we have considered the variance of one random variable at a time. However, the related concept of covariance measures the relationship between two random variables.

Definition 1.4.3. *The covariance of two random variables X and Y is given by*

$$Cov(X, Y) = E((X-\mu_X)(Y-\mu_Y))$$

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$.

Theorem 1.4.2. *Let X and Y be two random variables. Then*

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

Proof. Using linearity, we have

$$\begin{aligned} Cov(X, Y) =& E((X - \mu_X)(Y - \mu_Y)) = E(XY - X\mu_Y - \mu_X Y + \mu_X\mu_Y) \\ &= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y = E(XY) - 2\mu_X\mu_Y + \mu_X\mu_Y \\ Cov(X, Y) =& E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y) \end{aligned}$$

□

Exemple 1.4.4. *Let X and Y be discrete random variables, with joint probability function P_{XY} given*

$$P(X = x, Y = y) = \begin{cases} 1/2 & x = 3 =, y = 4 \\ 1/3 & x = 3, y = 6 \\ 1/6 & x = 5, y = 6 \\ 0 & otherwise \end{cases}$$

then $E(X) = 3(1/2 + 1/3) + 5(1/6) = 10/3$*,* $E(Y) = 4(1/2) + 6(1/3 + 1/6) = 5$*. and* $E(XY) = 3 \times 4(1/2) + 3 \times 6(1/3) + 5 \times 6(1/6) = 17$*. Hence*

$$Cov(X, Y) = E(XY) - E(X)E(Y) = 17 - (10/3)5 = 1/3$$

We note in particular that $Cov(X, Y)$ can be positive or a negative value. Intuitively, when positive, this says that Y increases when X increases, otherwise when negative, Y decreases when X increases.
We begin with some simple facts about covariance. Obviously, we always have $Cov(X, Y) = Cov(Y, X)$ We also have the following result.

Theorem 1.4.3. *(Linearity of covariance)*
Let X, Y, and Z be three random variables. Let a and b be real numbers. Then

$$Cov(aX + bY, Z) = aCov(X, Z) + bCov(Y, Z)$$

Corollary 1.4.2. *If X and Y are independent, then $Cov(X, Y) = 0$.*

Proof. Because X and Y are independent, we know (Theorem 1.3.3) that $E(XY) = E(X)E(Y)$. Hence, the result follows immediately from Theorem 1.4.2 □

We note that the converse to the previous Corollary is false, as the following example shows.

Exemple 1.4.5. *Covariance 0 Does Not Imply Independence.*
Let X and Y be discrete random variables, with joint probability function P_{XY} given by

$$P(X=x,Y=y)\begin{cases}1/4 & x=3,y=5\\ 1/4 & x=4,y=9\\ 1/4 & x=7,y=5\\ 1/4 & x=6,y=9\\ 0 & otherwise\end{cases}$$

then, $E(X)=3(1/4)+4(1/4)+7(1/4)+6(1/4)=5$*,* $E(Y)=5(1/4+1/4)+9(1/4+1/4)=7$ *and* $E(XY)=3\times 5(1/4)+4\times 9(1/4)+7\times 5(1/4)+6\times 9(1/4)=35$*, We obtain* $Cov(X,Y)=E(XY)-E(X)E(Y)=35-5\times 7=0$
On the other hand, X and Y are clearly not independent.

There is also an important relationship between variance and covariance.

Theorem 1.4.4. *For any random variables X and Y,*

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

Another concept very closely related to covariance is correlation.

Definition 1.4.4. *The correlation of two random variables X and Y is given by*

$$\rho(X,Y)=Corr(X,Y)=\frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

provided $Var(X),Var(Y)$ are not null.

Exemple 1.4.6. *Let X be any random variable with $Var(X)>0$, let $Y=3X$, and let $Z=-4X$. Then $Cov(X,Y)=3Var(X)$ and $Cov(X,Z)=-4Var(X)$ using theorem 1.4.4 . We know also that, $\sigma(Y)=3\sigma(X)$and $\sigma(Z)=4\sigma(X)$. Hence,*

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}=\frac{3Var(X)}{3\sigma(X)\sigma(X)}=1$$

because $\sigma(X)^2=Var(X)$. Also, we have that

$$\rho(X,Z)=\frac{Cov(X,Z)}{\sigma(X)\sigma(Z)}=\frac{-4Var(X)}{4\sigma(X)\sigma(X)}=-1$$

Intuitively, this again says that Y increases when X increases, whereas Z decreases when X increases. However, note that the scale factors 3 and (-4) have cancelled out, only their signs were important.

we always have $-1 \leq \rho(X, Y) \leq 1$, for any random variables X and Y. Hence, in the Example above, Y has the largest possible correlation with X (which makes sense because Y increases whenever X does, without exception), while Z has the smallest possible correlation with X (which makes sense because Z decreases whenever X does). We will also see that $\rho(X, Y)$ is a measure of the extent to which a linear relationship exists between X and Y.

SECTION 1.5

Conditional Expectation

We have seen before that conditioning on some event, or some random variable, can change various probabilities. Now, because expectations are defined in terms of probabilities, it seems reasonable that expectations should also change when conditioning on some event or random variable. Such modified expectations are called conditional expectations, as we now discuss.

1.5.1 Discrete case

The simplest case is when X is a discrete random variable, and A is some event of positive probability. We have the following.

Definition 1.5.1. *Let X be a discrete random variable, and let A be some event with $P(A) > 0$. Then the conditional expectation of X given A, is equal to*

$$E(X|A) = \sum_{x \in \mathbb{R}} x P(X = x|A) = \sum_{x \in \mathbb{R}} x \frac{P(X = x, A)}{P(A)}$$

Exemple 1.5.1. *Consider rolling a fair six-sided die, so that $S = \{1, 2, 3, 4, 5, 6\}$. Let X be the number showing, so that $X(s) = s$ for $s \in S$. Let $A = \{3, 5, 6\}$ be the event that the die shows 3, 5, or 6. What is $E(X|A)$?*
Here we know that
$P(X = s|A) = P(X = s|X = 3, 5or6)$ *and that,* $P(X = 3|A) = P(X = 3|X = 3, 5or6) = 1/3$ *and that, similarly,* $P(X = 5|A) = P(X = 6|A) = 1/3$*. While* $P(X = 1, 2, 4|A) = 0$*.*

Hence,

$$E(X|A) = 3P(X=3|A) + 5P(X=5|A) + 6P(X=6|A)$$
$$= 3(1/3) + 5(1/3) + 6(1/3) = 14/3$$

Often we wish to condition on the value of some other random variable. If the other random variable is also discrete, and if the conditioned value has positive probability, then this works as above.

Definition 1.5.2. *Let X and Y be discrete random variables, with $P(Y=y) > 0$. Then the conditional expectation of X given $Y=y$, is equal to*

$$E(X|Y=y) = \sum_{x\in\mathbb{R}} xP(X=x|Y=y) = \sum_{x\in\mathbb{R}} x\frac{P(X=x, Y=y)}{P(Y=y)}$$

Exemple 1.5.2. *Suppose the joint probability function of X and Y is given by*

$$P(X=x, Y=y) = \begin{cases} 1/7 & x=5, y=0 \\ 1/7 & x=5, y=3 \\ 1/7 & x=5, y=4 \\ 3/7 & x=8, y=0 \\ 1/7 & x=8, y=4 \\ 0 & otherwise \end{cases}$$

then,

$$E(X|Y=0) = \sum_{x\in\mathbb{R}} xP(X=x|Y=0)$$
$$= 5P(X=5|Y=0) + 8P(X=8|Y=0) = 5\frac{P(X=5, Y=0)}{P(Y=0)} + 8\frac{P(X=8, Y=0)}{P(Y=0)}$$
$$= 5\frac{1/7}{1/7+3/7)} + 8\frac{3/7}{1/7+3/7} = \frac{29}{4}$$

Similarly we find

$$E(X|Y=4) = \sum_{x\in\mathbb{R}} xP(X=x|Y=4)$$
$$= 5P(X=5|Y=4) + 8P(X=8|Y=4) = 5\frac{P(X=5, Y=4)}{P(Y=4)} + 8\frac{P(X=8, Y=4)}{P(Y=4)}$$
$$= 5\frac{1/7}{1/7+1/7)} + 8\frac{1/7}{1/7+1/7} = \frac{13}{2}$$

Also

$$E(X|Y=3) = \sum_{x\in\mathbb{R}} xP(X=x|Y=3)$$
$$= 5P(X=5|Y=3) = 5\,\frac{P(X=5,Y=3)}{P(Y=3)} = 5\,\frac{1/7}{1/7} = 5$$

Sometimes we wish to condition on a random variable Y, without specifying in advance on what value of Y we are conditioning. In this case, the conditional expectation $E(X|Y)$ is itself a random variable — namely, it depends on the (random) value of Y that occurs.

Definition 1.5.3. *Let X and Y be discrete random variables. Then the conditional expectation of X given Y, is the random variable $E(X|Y)$ which is equal to $E(X|Y=y)$ when $Y=y$. In particular, $E(X|Y)$ is a random variable that depends on the random value of Y*

Exemple 1.5.3. *Suppose again that the joint probability function of X and Y is given by*

$$P(X=x,Y=y) = \begin{cases} 1/7 & x=5, y=0 \\ 1/7 & x=5, y=3 \\ 1/7 & x=5, y=4 \\ 3/7 & x=8, y=0 \\ 1/7 & x=8, y=4 \\ 0 & otherwise \end{cases}$$

We have already computed that $E(X|Y=0) = 29/4, E(X|Y=4) = 13/2$, and $E(X|Y=3) = 5$. We can express these results together by saying that

$$E(X|Y) = \begin{cases} 29/4 & Y=0 \\ 5 & Y=3 \\ 13/2 & Y=4 \end{cases}$$

That is, $E(X|Y)$ is a random variable, which depends on the value of Y. Note that, because $P(Y=y)=0$ for $y \neq 0,3,4$, the random variable $E(X|Y)$ is undefined in that case.

Finally, we note that just like for regular expectation, conditional expectation is linear.

Theorem 1.5.1. *Let X_1, X_2, and Y be random variables, let A be an event, let ab, and y be real numbers, and let $Z = aX_1 + bX_2$. Then*
(a) $E(Z|A) = aE(X_1|A) + bE(X_2|A)$.
(b) $E(Z|Y=y) = aE(X_1|Y=y) + bE(X_2|Y=y)$.
(c) $E(Z|Y) = aE(X_1|Y) + bE(X_2|Y)$.

1.5.2 Absolutely Continuous Case

Definition 1.5.4. *Let X and Y be jointly absolutely continuous random variables, with joint density function $f(x,y)$. Then the conditional expectation of X given $Y = y$, is equal to*

$$E(X|Y=y) = \int_{\mathbb{R}} xf(x|y)dx = \int_{\mathbb{R}} x\frac{f(x,y)}{f(y)}dx$$

Exemple 1.5.4. *Let X and Y be jointly absolutely continuous, with joint density function $f(x,y)$ given by*

$$f(x,y) = \begin{cases} 4x^2y + 2y^5 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & otherwise \end{cases}$$

Then for $0 \leq y \leq 1$,

$$f(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^1 4x^2y + 2y^5 dx = 4y/3 + 2y^5$$

Hence,

$$E(X|Y=y) = \int_{-\infty}^{\infty} x\frac{f(x,y)}{f(y)}dx = \int_0^1 x\frac{4x^2y+2y^5}{4y/3+2y^5}dx = \frac{y+y^5}{4y/3+2y^5} = \frac{1+y^4}{4/3+2y^4}$$

As in the discrete case, we often wish to condition on a random variable without specifying in advance the value of that variable. Thus, $E(X|Y)$ is again a random variable, depending on the random value of Y.

Definition 1.5.5. *Let X and Y be jointly absolutely continuous random variables. Then the conditional expectation of X given Y, is the random variable $E(X|Y)$ which is equal to $E(X|Y=y)$ when $Y=y$. Thus, $E(X|Y)$ is a random variable that depends on the random value of Y.*

Exemple 1.5.5. *Let X and Y again have joint density*

$$f(x,y) = \begin{cases} 4x^2y + 2y^5 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & otherwise \end{cases}$$

We already know that $E(X|Y=y) = \frac{1+y^4}{4/3+2y^4}$

This formula is valid for any y between 0 and 1, so we conclude that

$$E(X|Y) = \frac{1 + y^4}{4/3 + 2y^4}$$

Note that in this last formula, Y is a random variable, so $E(X|Y)$ is also a random variable.

Finally, we note that in the absolutely continuous case, conditional expectation is still linear.

Chapter 2

Sampling Distributions and Limits

In many applications of probability theory, we will be faced with the following problem. Suppose that X_1, X_2, X_n is an identically and independently distributed (i.i.d.) sequence, i.e., X_1, X_2, X_n is a sample from some distribution, and we are interested in the distribution of a new random variable $Y = h(X_1, X_2, X_n)$ for some function h In particular, we might want to compute the distribution function of Y or perhaps its mean and variance. The distribution of Y is sometimes referred to as its sampling distribution, as Y is based on a sample from some underlying distribution.
Quite often, however, exact results are impossible to obtain, as the problem is just too complex. In such cases, we must develop an approximation to the distribution of Y
For many important problems, a version of Y is defined for each sample size n (e.g., a sample mean or sample variance), so that we can consider a sequence of random variables $Y_1, Y_2 ...$ etc. This leads us to consider the limiting distribution of such a sequence so that, when n is large, we can approximate the distribution of Y_n by the limit, which is often much simpler. This approach leads to a famous result, known as the central limit theorem, discussed below.

SECTION 2.1

Sampling Distributions

Let us consider a very simple example.

Exemple 2.1.1. *Suppose we obtain a sample X_1, X_2 of size $n = 2$ from the discrete distribution with probability function given by*

$$P_X(x) = \begin{cases} 1/2 & x = 1 \\ 1/4 & x = 2 \\ 1/4 & x = 3 \\ 0 & otherwise \end{cases}$$

Let us take $Y_2 = (X_1X_2)^{1/2}$. This is the geometric mean of the sample values (the geometric mean of n positive numbers $x_1, ..., x_n$ is defined as $(x_1...x_n)^{1/n}$. To determine the distribution of Y_2 we first list the possible values for Y_2 the samples that give rise to these values, and their probabilities of occurrence. The values of these probabilities specify the sampling distribution of Y We have the following table.

y	*Sample*	$P_{Y_2}(y)$
1	*(1,1)*	*(1/2)(1/2)= 1/4*
$\sqrt{2}$	*(1,2), (2,1)*	*(1/2)(1/4)+(1/4)(1/2)=1/4*
$\sqrt{3}$	*(1,3), (1,3)*	*(1/2)(1/4)+(1/4)(1/2)=1/4*
2	*(2,2)*	*(1/4)(1/4)= 1/16*
$\sqrt{6}$	*(2,3), (3,2)*	*(1/4)(1/4)+(1/4)(1/4)=1/8*
3	*(3,3)*	*(1/4)(1/4)= 1/16*

Now suppose instead we have a sample $X_1, ..., X_{20}$ of size $n = 20$ and we want to find the distribution of $Y_{20} = (X_1...X_{20})^{1/20}$. Obviously, we can proceed as above, but this time the computations are much more complicated, as there are now $3^{20} = 3,486,784,401$ possible samples, as opposed to the $3^2 = 9$ samples used to form the previous table. Directly computing $P_{Y_{20}}$, as we have done for P_{Y_2}, would be onerous — even for a computer! So what can we do here?
One possibility is to look at the distribution of $Y_n = (X_1...X_n)^{1/n}$ when n is large and see if we can approximate this in some fashion.

SECTION 2.2

Convergence in Probability

Notions of convergence are fundamental tool of mathematics. For example, if $a_n = 1 - 1/n$, then $a_1 = 0, a_2 = 1/2, a_3 = 2/3, a_4 = 3/4$, etc. We see that the values of a_n are getting " closer and closer " to 1, and indeed we know from calculus that $\lim_{n\to\infty} a_n = 1$ in this case. For random variables, notions of convergence are more complicated. If the values themselves are random, then how can they "converge" to anything? On the other hand, we can consider various probabilities associated with the random variables and see if they converge in some sense. The simplest notion of convergence of random variables is convergence in probability, as follows. (Other notions of convergence will be developed in subsequent sections.)

Definition 2.2.1. *Let $X_1, X_2...$ be an infinite sequence of random variables, and let Y be another random variable. Then the sequence X_n converges in probability to Y, if for all $\epsilon > 0$, we have*

$$\lim_{n\to\infty} P(|X_n - Y| \geq \epsilon) = 0$$

and we write $X_n \xrightarrow{P} Y$

In Figure 2.1, we have plotted the differences $(X_n - Y)$ for selected values of n for 10 generated sequences $\{X_n - Y\}$ for a typical situation where the random variables X_n converge to a random variable Y in probability We have also plotted the horizontal lines at for $\pm\epsilon$, $\epsilon = 0.25$ From this we can see the increasing concentration of the distribution of $X_n - Y$ about 0, as n increases, as required by Definition 2.2.1. In fact, the 10 observed values of $X_{100} - Y$ all satisfy the inequality $|X_{100} - Y| < 0.25$

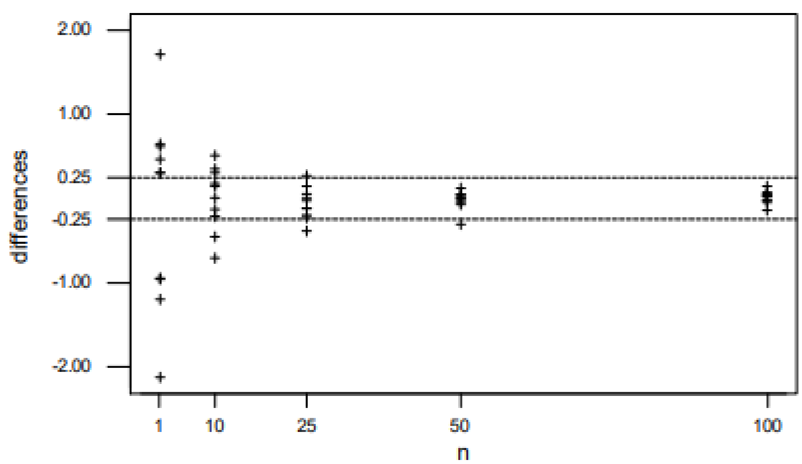


Figure 2.1: Plot of 10 replications of $X_n - Y$ illustrating the convergence in probability of X_n to Y

We consider some applications of this definition.

Exemple 2.2.1. *Let Y be any random variable, and let $X_1 = X_2 = X_3 = Y$. (That is, the random variables are all identical to each other.) In that case, $|X_n - Y| = 0$, so of course*

$$\lim_{n\to\infty} P(|X_n - Y| \geq \epsilon) = 0$$

for all $\epsilon > 0$. Hence, $X_n \xrightarrow{P} Y$.

Exemple 2.2.2. *Suppose $P(X_n = 1 + 1/n) = 1$ and $P(Y = 1) = 1$. Then $P(|X_n - Y| \geq \epsilon) = 1$ whenever $n < 1/\epsilon$. Hence, $P(|X_n - Y| \geq \epsilon) \to 0$ as $n \to \infty$ for all ϵ. Hence, the sequence X_n converges in probability to Y. (Here, the distributions of X_n and Y are all degenerate.)*

2.2.1 The Weak Law of Large Numbers

One of the most important applications of convergence in probability is the weak law of large numbers. Suppose that $X_1, X_2...$ is a sequence of independent random variables that each have the same mean. For large n, what can we say about their average

$$M_n = \frac{1}{n}(X_1 + \cdots + X_n)?$$

We refer to M_n as the sample average, or sample mean, for $X_1, ..., X_n$ When the sample size n is fixed, we will often use $\bar{X}$ as a notation for sample mean instead of M_n.
For example, if we flip a sequence of fair coins, and if $X_i = 1$ or $X_i = 0$ as the i^{th} coin comes up heads or tails, then M_n represents the fraction of the first n coins that came up heads. We might expect that for large n, this fraction will be close to $1/2$, i.e., to the expected value of the X_i.
The weak law of large numbers provides a precise sense in which average values M_n tend to get close to $E(X_i)$, for large n.

Theorem 2.2.1. *(Weak law of large numbers)*
Let $X_1, X_2...$ be a sequence of independent random variables, each having the same mean μ and each having variance less than or equal to $v < \infty$. Then,

$$\forall \epsilon \geq 0, \lim_{n\to\infty} P(|M_n - \mu| \geq \epsilon) = 0$$

That is, the averages converge in probability to the common mean μ ($M_n \xrightarrow{P} \mu$).

Exemple 2.2.3. *Consider flipping a sequence of identical coins, each of which has probability p of coming up heads. Let M_n again be the fraction of the first n coins that are heads. Then by the weak law of large numbers, for any $\epsilon \geq 0$, $\lim_{n\to\infty} P(p - \epsilon < M_n < p + \epsilon) = 1$. Thus we see that for a large n, it is very likely that M_n is very close to p. (The previous example corresponds to the special case $p = 1/2$.)*

Exemple 2.2.4. *Let $X_1, X_2...$ be i.i.d. with distribution $N(3, 5)$ Then by the weak law of large numbers, $P(3 - \epsilon < M_n < 3 + \epsilon) \to 1$ as $n \to \infty$. Hence, for a larger n, the average value M_n is very close to 3.*

SECTION 2.3

Convergence Almost Surely

A notion of convergence for random variables that is closely associated with the convergence of a sequence of real numbers is provided by the concept of convergence almost surely (with probability 1). This property is given in the following definition.

Definition 2.3.1. *Let $X_1, X_2...$ be an infinite sequence of random variables. We shall say that the sequence X_i converges almost surely (a.s.) (with probability 1) to a random variable Y, if $P(\lim_{n\to\infty} X_n = Y) = 1$ and we write $X_n \overset{a.s}{\to} Y$.*

In Figure 2.2, we illustrate this convergence by graphing the sequence of differences $(X_n - Y)$ for a typical situation where the random variables X_n converge to a random variable Y almost surely. We have also plotted the horizontal lines at $\pm\epsilon$ for $\epsilon = 0.1$. Notice that inevitably all the values $(X_n - Y)$ are in the interval (-0,1 0.1) or, in other words, the values of X_n are within 0.1 of the values of Y
Definition 2.3.1 indicates that for any given $\epsilon \geq 0$ there will exist a value n_ϵ such that $(X_n - Y)$ for every $n \geq n_\epsilon$ The value of n_ϵ will vary depending on the observed value of the sequence $(X_n - Y)$ but it always exists. Contrast this with the situation depicted in Figure 2.1, which only says that the probability distribution $(X_n - Y)$ concentrates about 0 as n grows and not that the individual values of $(X_n - Y)$ will necessarily all be near 0.

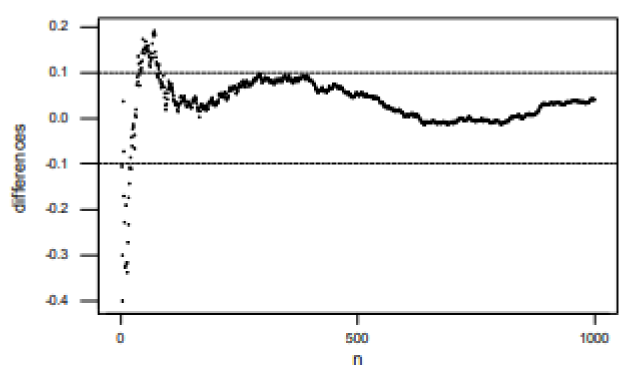


Figure 2.2: Plot of a single replication $\{X_n - Y\}$ illustrating the convergence almost surely of X_n to Y.

One might wonder what the relationship is between convergence in probability and convergence almost surely. The following theorem provides an answer.

Theorem 2.3.1. *Let $Z, Z_1, Z_2...$ be random variables. Suppose Z_n converges to Z almost surely. Then Z_n converges to Z in probability. That is, if a sequence of random variables converges almost surely, then it converges in probability to the same limit.*

Lemma 2.3.1. *Let $X_1, X_2, ...$ be a sequence of events in some probability space. The Borel–Cantelli lemma states If the sum of the probabilities of the events $\{X_n\}$ is finite*

$$\sum_{n\geq 1} P(X_n) < \infty$$

then the probability that infinitely many of them occur is 0, that is,

$$P(\lim_{n\to\infty} supX_n) = 0$$

A related result, sometimes called the second Borel–Cantelli lemma, states If the events (X_n) are independent and the sum of the probabilities of the (X_n) diverges to infinity, then the probability that infinitely many of them occur is 1. That is

If $\sum_{n\geq 1} P(X_n) = \infty$ and the events $(X_n)_{n=1}^{\infty}$ are independent, then $P(\lim_{n\to\infty} supX_n) = 1$

On the other hand, the converse to the Theorem above is false, as the following example shows.

Exemple 2.3.1. *Consider a sequence (X_n) of independent random variables such that $P(X_n = 1) = \frac{1}{n}$ and $P(X_n = 0) = 1 - \frac{1}{n}$.*
For $0 < \epsilon < 1/2$, we have $P(|X_n| \geq \epsilon) = \frac{1}{n}$ which converges to 0, hence $X_n \xrightarrow{P} 0$.
Since $\sum_{n\geq 1} P(X_n = 1) = +\infty$ and the events $(X_n = 1)$ are independent, second Borel-Cantelli lemma ensures that $P(\lim_{n\to\infty} supXn = 1) = 1$.
Hence the sequence (X_n) does not converge to 0 almost everywhere (in fact the set on which this sequence does not converge to 0 has probability 1).

2.3.1 The Strong Law of Large Numbers

The following is a strengthening of the weak law of large numbers because it concludes convergence with almost surely probability instead of just convergence in probability.

Theorem 2.3.2. *(Strong law of large numbers)*
Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables, each having finite mean. Then

$$P(\lim_{n\to\infty} M_n = \mu) = 1$$

That is, the averages converge almost surely to the common mean or $M_n \stackrel{a.s}{\to} \mu$.

This result says that sample averages converge almost surely to μ. Like Theorem 2.2.1, it says that for larger n the averages M_n are usually close to $E(X_i)$. But it says in addition that if we wait long enough (i.e., if n is large enough), then eventually the averages will all be close to μ. In other words, the sample mean is consistent for μ

SECTION 2.4

Convergence in Distribution

There is yet another notion of convergence of a sequence of random variables that is important in applications of probability and statistics.

Definition 2.4.1. *Let $X, X_1, X_2, ...$ be random variables. Then we say that the sequence (X_n) converges in distribution to X if for all $x \in \mathbb{R}$ $F_{X_n}(x) \to F_X(x)$ where $F(.)$ is the cumulative distribution function, i.e.*

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x)$$

and we write $X_n \stackrel{D}{\to} X$

Intuitively, (X_n) converges in distribution to X if for large n, the distribution of X_n is close to that of X. The importance of this, as we will see, is that often the distribution of X_n is difficult to work with, while that of X is much simpler. With X_n converging in distribution to X however, we can approximate the distribution of X_n by that of X.

Exemple 2.4.1. *Suppose $P(X_n = 1) = 1/n$, and $P(X_n = 0) = 1 - 1/n$. Let $X = 0$ so that $P(X = 0) = 1$. Then,*

$$P(X_n \leq x) = \begin{cases} 0 & x < 0 \\ 1 - 1/n & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}, P(X = x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x \end{cases}$$

as $n \to \infty$ As $P(X_n \leq x) \to P(X \leq x)$ for every x and in particular at all x where $P(X = x) = 0$ we have that (X_n) converges in distribution to X. Intuitively, as $n \to \infty$, it is more and more likely that X_n will equal 0.

Exemple 2.4.2. *Suppose $P(X_n = 1) = 1/2 + 1/n$, and $P(X_n = 0) = 1/2 - 1/n$. Suppose further that $P(X = 0) = P(X = 1) = 1/2$. Then X_n converges in distribution to X because $P(X_n = 1) = 1/2$ and $P(X_n = 0) = 1/2$ as $n \to \infty$.*

Exemple 2.4.3. *Poisson Approximation of the Binomial distribution*
Suppose $X_n \rightsquigarrow$ Binomial$(n, \lambda/n)$ and $X \rightsquigarrow$ Poisson(λ) . We have seen before that

$$P(X_n = k) = C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \overset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!}$$

This implies that $F_{X_n}(x) \to F_X(x)$ Therefore, X_n converges in distribution to X. (Indeed, this was our original motivation for the Poisson distribution.)

More examples of convergence in distribution are given by the central limit theorem, discussed in the next section. We first pause to consider the relationship of convergence in distribution to our previous notions of convergence.

Theorem 2.4.1. *If $X_n \overset{P}{\to} X$, then $X_n \overset{D}{\to} X$*

The converse to this theorem is false. Indeed, the fact that X_n converges in distribution to X says nothing about the underlying relationship between X_n and X, it says only something about their distributions. The following example illustrates this.

Exemple 2.4.4. *Suppose $X, X_1, X_2, ...$ are i.i.d., each equal to ∓ 1 with probability 1/2 each. In this case, $P(X_n \leq x) = P(X \leq x)$ for all n and for all $x \in \mathbb{R}$, so of course X_n converges in distribution to X. On the other hand, because X and X_n are independent,*

$$P(|X_n - X| \geq 2) = 1/2$$

for all n, which does not go to 0 as $n \to \infty$. Hence, X_n does not converge to X in probability. So we can have convergence in distribution without having convergence in probability.

Finally, we note that combining Theorem 2.3.1 with Theorem 2.4.1 reveals the following.

Corollary 2.4.1. *If X_n converge to X almost surely, then $X_n \overset{D}{\to} X$*

2.4.1 The Central Limit Theorem

We now present the central limit theorem, one of the most important results in all of probability theory. Intuitively, it says that a large sum of i.i.d. random variables, properly normalized, will always have approximately a normal distribution. This shows that the normal distribution is extremely fundamental in probability and statistics even though its density function is complicated and its cumulative distribution function is intractable.
Suppose $X_1, X_2, ...$ is an i.i.d. sequence of random variables each having finite mean μ and finite variance σ^2 Let $S_n = X_1 + \cdots + X_n$ be the sample sum and $M_n = S_n/n$ be the sample mean. The central limit theorem is concerned with the distribution of the random variable

$$Z_n \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{M_n - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\frac{M_n - \mu}{\sigma}\right)$$

where $\sigma = \sqrt{\sigma^2}$ We know $E(M_n) = \mu$ and $Var(M_n) = \sigma^2/\sqrt{n}$ which implies that $E(Z_n) = 0$ and $Var(Z_n) = 1$ The variable Z_n is thus obtained from the sample mean (or sample sum) by subtracting its mean and dividing by its standard deviation. This transformation is referred to as standardizing a random variable, so that it has mean 0 and variance 1. Therefore, Z_n is the standardized version of the sample mean (sample sum.
Note that the distribution of Z_n shares two characteristics with the $N(0, 1)$ distribution, namely, it has mean 0 and variance 1. The central limit theorem shows that there is an even stronger relationship.

Theorem 2.4.2. *(The central limit theorem)*
Let $X_1, X_2, ...$ be i.i.d. with finite mean μ and finite variance σ^2. Let $Z \rightsquigarrow N(0, 1)$. Then as $n \to \infty$, the sequence Z_n converges in distribution to Z, i.e., $Z_n \xrightarrow{D} Z$.

The central limit theorem is so important that we shall restate its conclusions in several different ways.

Corollary 2.4.2. *For each fixed $x \in \mathbb{R}$, $\lim_{n\to\infty} P(Z_n \le x) = \Phi(x)$ where $\Phi(x)$ is the cumulative distribution function for the standard normal distribution.*

We can write this as follows.

Corollary 2.4.3. *For each fixed $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} P(S_n \le n\mu + x\sqrt{(}n)\sigma) = \Phi(x), \ \lim_{n\to\infty} P(M_n \le \mu + x\sigma)/sqrtn) = \Phi(x)$$

We note that it is not essential in the central limit theorem to divide by σ, in which case the theorem asserts instead that $(S_n - n\mu)/\sqrt{n}$ (or $\sqrt{n}(M_n - \mu)$) converges in distribution to the $N(0, \sigma^2)$ distribution. That is, the limiting distribution will still be normal but will have variance σ^2 instead of variance 1.
Similarly, instead of dividing by exactly σ , it suffices to divide by any quantity σ_n, provided $\sigma_n \overset{a.s}{\to} \sigma$.

Corollary 2.4.4. *If we have the variable Z_n^*, when*

$$Z_n^* \frac{S_n - n\mu}{\sqrt{n}\sigma_n} = \frac{M_n - \mu}{\sigma_n/\sqrt{n}} = \sqrt{n}\left(\frac{M_n - \mu}{\sigma_n}\right)$$

and $\sigma_n \overset{a.s}{\to} \sigma$, then $Z_n^* \overset{D}{\to} Z$ as $n \to \infty$.

To illustrate the central limit theorem, we consider a simulation experiment.

Exemple 2.4.5. *(The Central Limit Theorem Illustrated in a Simulation)*
Suppose we generate a sample $X_1, ..., X_n$ from the Uniform$[0, 1]$ density. Note that the Uniform$[0, 1]$ density is completely unlike a normal density. An easy calculation shows that when $X \rightsquigarrow$ Uniform$[0, 1]$ then $E(X) = 1/2$ and $Var(X) = 1/12$.
Now suppose we are interested in the distribution of the sample average $M_n = S_n/n = (X_1 + \cdots + X_n)/n$ for various choices of n The central limit theorem tells us that

$$Z_n = \frac{S_n - n/2}{\sqrt{n/12}} = \sqrt{n}\left(\frac{M_n - 1/2}{\sqrt{1/12}}\right)$$

converges in distribution to an $N(0, 1)$ distribution.

Exemple 2.4.6. *For example, suppose $X_1, X_2, ...$ are i.i.d. random variables, each with the Poisson(5) distribution. Recall that this implies that $E(X_i) = 5$and 2 $Var(X_i) = 5$.Hence, for each fixed $x \in \mathbb{R}$, we have*

$$P(S_n \leq 5n + x\sqrt{5n}) \to \Phi(x)$$

as $n \to \infty$

Exemple 2.4.7. *For example, suppose that we have a biased coin, where the probability of getting a head on a single toss is $\theta = 0.6$ We will toss the coin $n = 1000$ times and then calculate the probability of getting at least 550 heads and no more than 625 heads. If Y denotes the number of heads obtained in the 1000 tosses, we have that $Y \rightsquigarrow$Binomial$(1000, 0.6)$ so*

$$E(Y) = 1000(0.6) = 600$$
$$Var(Y) = 1000(0.6)(0.4) = 240$$

Note that we are approximating a discrete distribution by a continuous distribution here. Reflecting this, a small improvement (which is called the correction for continuity) is often made, by adding 0.5, i.e., we consider the interval $(y - 0.5, y + 0.5)$ to the non-negative integer y. Therefore,

$$\begin{aligned}
P(550 \leq Y \leq 625) =& P(550 + 0.5 \leq Y \leq 625 + 0.5) \\
&= P(\frac{549.5 - 600}{\sqrt{240}} \leq \frac{Y - 600}{\sqrt{240}} \leq \frac{625.5 - 600}{\sqrt{240}} \\
&= P(-3.2598 \leq \frac{Y - 600}{\sqrt{240}} \leq 1.646) \\
&= \Phi(1.65) - \Phi(3.26)
\end{aligned}$$

Note that it would be impossible to compute this probability using the formulas for the binomial distribution.

2.4.2 Cumulative Distribution Function of the Standard Normal Distribution

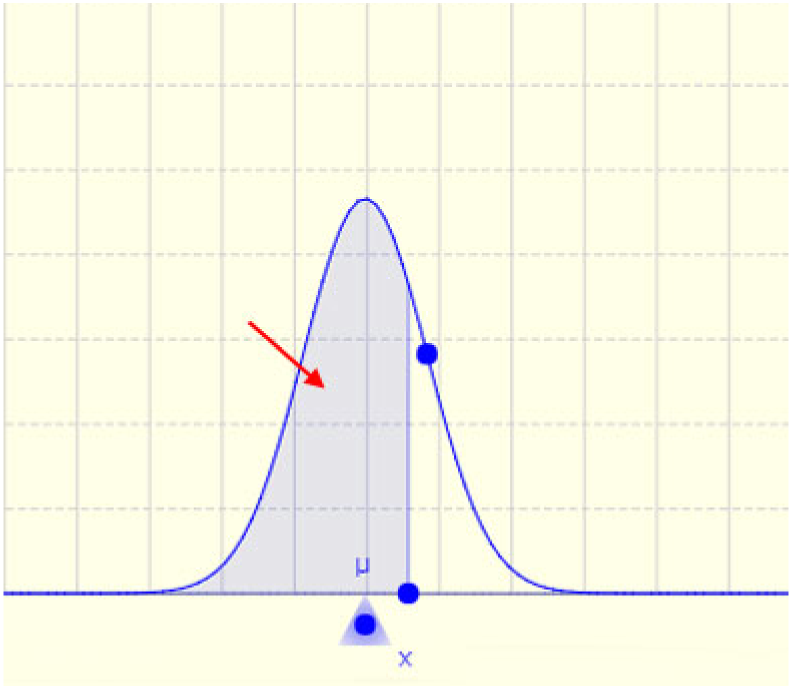
The (cumulative) distribution function of a random variable X, evaluated at x, is the probability that X will take a value less than or equal to x.

$$F(x) = P(X \leq x)$$

In the case of a continuous distribution (like the normal distribution) it is the area under the probability density function (the 'bell curve') from the negative left (minus infinity) to x. The shaded area of the curve represents the probability that X is less or equal than x.

We can use the integral notation, then the (cumulative) distribution function can be written as an integral of its probability density function:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$



In the case of the normal distribution this integral does not exist in a simple closed formula. It is computed numerically.
The Standard normal distribution plays an important role

$$Z \rightsquigarrow N(0,1)$$

In some books they use a special notation for the (cumulative) distribution function in this special case of a standard normal distribution:

$$\Phi(z)$$

We already know that there is a relation between any normal distribution X and the standard normal distribution Z with mean 0 and standard deviation 1, so

$$F(x) = \Phi(\frac{x-\mu}{\sigma})$$

the following table represents the standard normal CFD values,

	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

This can be clarified by a few simple examples.
What is the probability that Z is less than or equal to 1.53? ($\Phi(1.53)$) Look for 1.5 in the first column, go right to the 0.03 column to find the value 0.9370
What is the probability that Z is less than or equal to -1.53? For negative values, use the relationship

$$\Phi(-a) = 1 - \Phi(a)$$

From the first example, this gives 1 - 0.9370 = 0.0630.
What is the probability that Z is between a and b, for this case we use

$$P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$$

SECTION 2.5

Monte Carlo Approximations

According to the laws of large numbers if $X_1, X_2...$ is an i.i.d. sequence of random variables with mean μ, and

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

,

then for large n we will have $M_n \approx \mu$.

Suppose now that μ is unknown. Then, as discussed in Section 4.4.2, it is possible to change perspective and use M_n (for large n) as an estimator or approximation of μ.

Exemple 2.5.1. *Suppose we believe a certain medicine lowers blood pressure, but we do not know by how much. We would like to know the mean amount , by which this medicine lowers blood pressure. Suppose we observe n patients (chosen at random so they are i.i.d.), where patient i has blood pressure B_i before taking the medicine and blood pressure A_i afterwards. Let $X_i = B_i A_i$ Then*

$$M_n = \frac{1}{n}\sum_{i=1}^{n}(B_i - A_i)$$

is the average amount of blood pressure decrease. (Note that $B_i - A_i$ may be negative for some patients, and it is important to also include those negative terms in the sum.) Then for large n, the value of M_n is a good estimate of $E(X_i)$.

Such estimators can also be used to estimate purely mathematical quantities that do not involve any experimental data (such as coins or medical patients) but that are too difficult to compute directly. In this case, such estimators are called Monte Carlo approximations (named after the gambling casino in the principality of Monaco because they introduce randomness to solve non-random problems).

Exemple 2.5.2. *Suppose we wish to evaluate*

$$I = \int_0^1 \cos(x)^2 \sin(x)^4 dx$$

This integral cannot easily be solved exactly. But it can be approximately computed using a Monte Carlo approximation, as follows. We note that

$$I = E(\cos(U)^2 \sin(U)^4$$

where $U \rightsquigarrow$ Uniform$[0,1]$. Hence, for large n, the integral I is approximately equal to $M_n = \frac{T_1+\cdots+T_n}{n}$, where $T_i = \cos(U_i)^2\sin(U_i)^4$, and where $U_1, U_2, ...$ are i.i.d. Uniform$[0,1]$.

Chapter 3

Statistical Inference

In this chapter, we begin our discussion of statistical inference. Probability theory is primarily concerned with calculating various quantities associated with a probability model. This requires that we know what the correct probability model is. In applications, this is often not the case, and the best we can say is that the correct probability measure to use is in a set of possible probability measures. We refer to this collection as the statistical model. So, in a sense, our uncertainty has increased, not only do we have the uncertainty associated with an outcome or response as described by a probability measure, but now we are also uncertain about what the probability measure is. Statistical inference is concerned with making statements or inferences about characteristics of the true underlying probability measure. Of course, these inferences must be based on some kind of information, the statistical model makes up part of it. Another important part of the information will be given by an observed outcome or response, which we refer to as the data. Inferences then take the form of various statements about the true underlying probability measure from which the data were obtained.
Inference includes estimation and hypothesis testing which are discussed in this chapter and the next one.

SECTION 3.1

Statistical model

In a statistical context, we observe the data s, but we are uncertain about P. In such a situation, we want to construct inferences about P based on this data.
How we should go about making these statistical inferences is probably not at all obvious. Common to virtually all approaches to statistical inference is the concept of the statistical model for the data. This takes the form of a set $\{P_\theta, \theta \in \Theta :\}$ of probability measures, one

of which corresponds to the true unknown probability measure P that produced the data s In other words, we are asserting that there is a random mechanism generating s and we know that the corresponding probability measure P is one of the probability measures in $\{P_\theta, \theta \in \Theta :\}$.
The statistical model $\{P_\theta, \theta \in \Theta :\}$: corresponds to the information a statistician brings to the application about what the true probability measure is, or at least what one is willing to assume about it. The variable θ is called the parameter of the model, and the set Θ is called the parameter space. Typically, we use models where $\theta \in \Theta$ indexes the probability measures in the model, i.e., $P_{\theta_1} = P_{\theta_2}$ if and only if $\theta_1 = \theta_2$ If the probability measures P_θ can all be presented via probability functions or density functions f_θ
From the definition of a statistical model, we see that there is a unique value $\theta \in \Theta$ such that P_θ is the true probability measure. We refer to this value as the true parameter value. It is obviously equivalent to talk about making inferences about the true parameter value rather than the true probability measure, i.e., an inference about the true value of θ is at once an inference about the true probability distribution. So, for example, we may wish to estimate the true value of θ.
There are two main approaches to estimation; point estimation and confidence interval estimation. The first one gives an approximate value for the unknown parameter, while the second gives a an interval that likely contains the value of the parameter.

SECTION 3.2

Point estimation

Point estimation method is based on the notion of estimators, this notion is defined by the following concepts.

Definition 3.2.1. *We call a statistic any function of the data in a sample* $(X_1, ..., X_n)$. *which is denoted by* $T_n(X_1, ..., X_n)$. *A statistic does not depend on unknown parameters.*

Definition 3.2.2. *Let* X *be a random variable whose distribution depends on a parameter* θ, *and let* $X_1, X_2, \ldots, X_n$ *be a size* n *sampling of* X. *A point estimator of* θ *is a statistic of the form*

$$\hat{\theta} = T(X_1, X_2, \ldots, X_n).$$

In an application, we want to know how reliable an estimator $\hat{\theta}$ is. or we might have to choose between two estimators of the same parameter. This leads us to following concepts.

Definition 3.2.3. *The bias in an estimator* $\hat{\theta}$ *of* θ *is given by* $E(\hat{\theta}) - \theta$ *whenever* $E(\hat{\theta})$ *exists. When the bias in an estimator* $\hat{\theta}$ *is 0, we call* $\hat{\theta}$ *an unbiased estimator of* θ*, i.e., T is unbiased whenever* $E(\hat{\theta}) = \theta$.

Unbiasedness tells us that, in a sense, the sampling distribution of the estimator is centered on the true value θ.

Definition 3.2.4. *The mean squared error (MSE) of the estimator $\hat{\theta}$ is given by $MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2)$.*

Clearly, the smaller $MSE(\hat{\theta})$ is, the more concentrated the sampling distribution of $\hat{\theta}$ is about the value θ. Looking at $MSE(\hat{\theta})$ as a function of gives us some idea of how reliable T is as an estimate of the true value of θ.
The following result gives an important identity for the MSE.

Theorem 3.2.1. *The mean squared error (MSE) is also expressed by*

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

Note that when the bias in an estimator is 0, then the MSE is just the variance and

$$Sd(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$$

is an estimate of the standard deviation of $\hat{\theta}$ and is referred to as the standard error of the estimate. As a principle of good statistical practice, whenever we quote an estimate of a quantity, we should also provide its standard error at least when we have an unbiased estimator, as this tells us something about the accuracy of the estimate.

Definition 3.2.5. *An estimator is said to be convergent (consistent) if the sequence $(\hat{\theta}_n)$ converges in probability to θ :*

$$\forall \epsilon > 0, P(|\hat{\theta}_n - \theta| > \epsilon) \underset{n\to\infty}{\longrightarrow} 0.$$

A Strong convergent (consistent) estimator is one where convergence is almost sure.
If the variance of an estimator tends to zero then, this estimator is consistent. this condition is sufficient not necessary.

There are several methods for determining point estimators, the method of moments, the maximum likelihood method and other methods.

3.2.1 Method of moments

In short, the method of moments involves equating sample moments with theoretical moments. So, let's start by making sure we recall the definitions of theoretical moments, as well as learn the definitions of sample moments.

Definition 3.2.6. *Moments(Review)*
Let X be a random variable. Then, The k^{th} moment of X is:

$$E(X^k)$$

and the k^{th} central moment of X is:

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

Usually, we are interested in the first moment of X, $\mu = E(X)$, and the second central moment of X, $Var(X) = E[(X - \mu)^2]$.

Definition 3.2.7. *Sample moments*
Let X be a random variable. Let $X_1, ..., X_n$ be iid realizations (samples) from X. Then, The k^{th} sample moment of X is:

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

and the k^{th} central sample moment of X is:

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

where $\bar{X}$ is the first sample moment.
For example, the first sample moment is just the sample mean, and the second central sample moment is the sample variance.

Common estimators are the sample mean and sample variance which are used to estimate the unknown population mean and variance.

Theorem 3.2.2. *(Estimation of μ)*
Suppose that the mean μ is unknown. The method of moments estimator of μ based on (X_n) is the sample mean

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

$\hat{\mu}$ is unbiased and consistent estimator.

Proof. $\bar{X}$ is unbiased because :
$E(\bar{X}) = E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = E(X_i) = \mu$.

$\bar{X}$ is consistent because :
$Var((\hat{\mu}) = Var(\bar{X}) = \sigma^2/n.$

$\square$

Estimating the variance of the distribution, on the other hand, depends on whether the distribution mean μ is known or unknown. First, we will consider the case when the mean in known.

Theorem 3.2.3. *Suppose that the mean μ is known and the variance σ^2 unknown, the method of moments estimator of σ^2 based on (X_n) is*

$$T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

T is is unbiased and consistent estimator.

Proof. T is unbiased because :
$E(T) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right) = \frac{1}{n}\sum_{i=1}^{n} Var(X_i) = \sigma^2.$

T is consistent because :
$Var(T) = Var\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right)$
$= \frac{1}{n^2}\sum_{i=1}^{n} Var[(X_i - \mu)^2]^2 = \frac{1}{n}\left[E((X_i - \mu)^4) - E((X_i - \mu)^2)^2\right].$

$\square$

Secondly, we will consider the more realistic case when the mean in also unknown.

Theorem 3.2.4. *Suppose that the mean μ and the variance σ^2 are both unknown, the method of moments estimator of σ^2 based on X_n is*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

S^2 is biased and consistent estimator.

Proof.

$$E(S^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{1}{n}\sum_{i=1}^{n}E(X_i - \bar{X})^2$$
$$= \frac{1}{n}\sum_{i=1}^{n}E(X_i^2) - E(\bar{X}^2)$$
$$= \frac{1}{n}(n(\mu^2 + \sigma^2)) - (\mu^2 + \frac{\sigma^2}{n}) = \frac{n-1}{n}\sigma^2$$

since $E(S^2) \to \sigma^2$ we call S^2 asymptotically unbiased.

$$Var(S^2) = \frac{1}{n}(\sigma_4 - \sigma^4) - \frac{2}{n^2}(\sigma_4 - 2\sigma^4) + \frac{1}{n^3}(\sigma_4 - 3\sigma^4)$$

where $\sigma_4 = E((X-\mu)^4)$, so S^2 is consistent. □

To obtain unbiased estimator of σ^2 while μ is unknown we consider the following estimator

Theorem 3.2.5.

$$S^{*2} = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

is an unbiased and consistent estimator of σ^2

Proof.

$$E(S^{*2}) = E\left(\frac{n}{n-1}S^2\right) = \frac{n}{n-1}E(S^2) = \sigma^2,$$

$$Var(S^{*2}) = \frac{1}{n}(\sigma_4 - \frac{n-3}{n-1}\sigma^4),$$

so so S^{*2} is unbiased and consistent.

□

Exemple 3.2.1. *Let* $X_1, X_2, ..., X_n$ *be (iid) Bernoulli random variables with parameter* θ*. We find the method of moments estimator of* θ*, the first theoretical moment is* $E(X_i) = \theta$ *We have just one parameter for which we are trying to derive the method of moments estimator. Therefore, we need just one equation. Equating the first theoretical moment with the corresponding sample moment, we get:*

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

Now, we just have to solve for θ. In this case, the equation is already solved, so

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

Exemple 3.2.2. *Let $X_1, X_2, ..., X_n$ be iid samples from $X \rightsquigarrow Exp(\lambda)$. The method of moment estimator of λ). We then set the first true moment to the first sample moment as follows (recall that $E(X) = \frac{1}{\lambda}$*

$$E[X] = \frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Solving for λ (just taking inverse), we get:

$$\hat{\lambda}_M = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}$$

3.2.2 Maximum of likelihood estimation

Given a parameter θ of a population to be estimated. We introduce the following estimator.

Definition 3.2.8. *Given $x = (x_1, x_2, \ldots, x_n)$ a realization of a sample $X = (X_1, X_2, \ldots, X_n)$ of n random variables, the function $L(\theta, x)$ is given by*

$$L(\theta, x) = \begin{cases} \prod_{i=1}^{n} P(X_i, \theta) & \text{if X is discrete} \\ \prod_{i=1}^{n} f(x_i, \theta) & \text{if X is continuous} \end{cases}$$

such that the function L of θ for x fixed, is called the Likelihood function.
The method of maximum likelihood (ML) consists in choosing as an estimator of θ, the particular value of θ which maximises the likelihood function $L(\theta, x)$.
this estimator $\hat{\theta}_{ML}$ is the solution to the equation:

$$\frac{\partial L(\theta, x)}{\partial \theta} = 0$$

or

$$\frac{\partial l(\theta, x)}{\partial \theta} = 0$$

where $l(\theta, x) = \ln L(\theta, x)$.

Exemple 3.2.3. *Suppose that $X = (X_1, X_2, ..., X_n)$ represents the outcomes of n independent Bernoulli trials, each with success probability θ. The likelihood for θ based on X is defined as the joint probability distribution of $X = (X_1, X_2, ..., X_n)$. Since $X = (X_1, X_2, ..., X_n)$ are iid random variables, the joint distribution is*

$$\begin{aligned} L(\theta, x) &= \prod_{i=1}^{n} P(x_i, \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{x_1}(1-\theta)^{1-x_1} \cdots \theta^{x_n}(1-\theta)^{1-x_n} \\ &= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

just reflects the probability mass function of the Bernoulli distribution.
For the log-likelihood,

$$l(\theta, x) = \ln L(\theta, x) = \left(\sum_{i=1}^{n} x_i\right) \ln\theta + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-\theta) = n\bar{X}\ln\theta + n(1-\bar{X})\ln(1-\theta)$$

Differentiating and setting it up to zero,

$$\frac{\partial l(\theta, x)}{\partial\theta} = n\left(\frac{\bar{X}}{\theta} - \frac{1-\bar{X}}{1-\theta}\right) = n\left(\frac{\bar{X}(1-\theta) - \theta(1-\bar{X})}{\theta(1-\theta)}\right) = 0$$

Therefore $\hat{\theta} = \bar{X}$

Exemple 3.2.4. *Let $X_1, \ldots, X_n$ be a random sample of size n drawn from a population distributed according to an exponential law Exp(λ), i.e.*

$$f(x, \lambda = \begin{cases} \lambda\exp(-\lambda x) & if x \geq 0 \\ 0 & if x < 0 \end{cases}$$

Estimate the parameter λ of the distribution, using the likelihood method.

$$\begin{aligned} L(\lambda, x) &= \prod_{i=1}^{n} f(x_i, \lambda) = \lambda\exp(-\lambda x_1) \times \lambda\exp(-\lambda x_2) \times \cdots \times \lambda\exp(-\lambda x_n) \\ &= \lambda^n \exp(-\lambda\sum_{i=1}^{n} x_i) \end{aligned}$$

$$l(\lambda, x) = \ln L(\lambda) = n\ln\lambda - \lambda\sum_{i=1}^{n} x_i$$

and $\dfrac{\partial}{\partial\lambda} l(\lambda, x) = \dfrac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \lambda = \dfrac{n}{\sum_{i=1}^{n} x_i}$
or we deduce that the M.V. estimator of λ *is* $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} x_i} = \dfrac{1}{\bar{X}}$.

SECTION 3.3

Confidence interval estimation

Let $\alpha \in]0, 1[$ be a risk level set by the statistician, A confidence region of θ of confidence level $1 - \alpha$ is a set (depending on on the observation but not on the unknown parameter θ), $C(X) \subseteq \Theta$, such that

$$\forall \theta \in \Theta, P(\theta \in C(X)) = 1 - \alpha$$

The usual values of α are 1%, 5% or 10%. In the one-dimensional case, most of the time, a confidence region is written in the form of an interval. A confidence interval with a 95% confidence level has a probability of at least 0.95 of containing the true unknown value θ. The estimation by interval of an unknown parameter θ is the construction of an interval $[a, b]$., we have :

$$P([a \leq \theta \leq b]) = 1 - \alpha$$

such that a et b called confidence limits.
in the following section, we consider X a random variable with normal distribution $N(\mu, \sigma^2)$., otherwise using the central limit theorem we get the same results when n is large $(n \geq 30)$.

3.3.1 Classic examples of interval estimation

The parameters to be estimated are the mean μ and the variance σ^2.
Starting with the mean we can distinguish two cases can be distinguished, depending on whether the variance is known or estimated.

Confidence interval of the mean

The unbiased estimator of the mean μ is the statistic

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

which follows the normal distribution $N((\mu, \frac{\sigma^2}{n})$,

Case 1: the variance σ^2 is known

Given a risk level α, we construct a probability interval for the sample mean $\bar{X}$ a probability interval:

$$P(-Z_{\alpha/2} \leq \sqrt{n}\frac{\bar{X}-\mu}{\sigma} \leq Z_{\alpha/2}) = 1-\alpha$$

$$P(-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P(-\bar{X}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P(\bar{X}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

Knowing that $\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \rightsquigarrow N(0,1)$: the value $Z_{\alpha/2}$ is read from the standard normal table $N(0,1)$ such that $\Phi(Z_{\alpha/2}) = 1-\dfrac{\alpha}{2}$. i.e., $Z_{\alpha/2} = \Phi^{-1}(1-\frac{\alpha}{2})$.

$$IC_\alpha(\mu) = \left[\bar{X}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Exemple 3.3.1. *The mass X of containers of a certain product is a random variable with mean μ and standard deviation 0.3 grams. A random sample of 100 containers is chosen and gives a mean of 49.7 grams.*
- Compute a confidence interval for μ with a confidence level of $1-\alpha = 0.95$.
Ans: Since the Variance σ^2 is known and the size n is large, the interval is of the form

$$\mu \in \bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$1-\alpha = 0.95 \Rightarrow Z_{\alpha/2} = Z_{0.025} = 1.96$. *where :*
$\mu \in 49.7 \pm 1.96 \times \dfrac{0.3}{\sqrt{100}} = [49.64, 49.76]$

Case 2 : the variance σ^2 is estimated

The unbiased estimator of the mean σ^2 is

$$S^{*2} = \frac{n}{n-1}\bar{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

When σ^2 is unknown, a confidence interval at the $1-\alpha$ level of μ the same as before while replacing σ of its estimate.

$$IC_\alpha(\mu) = \left[\bar{X} - Z_{\alpha/2}\frac{S^*}{\sqrt{n}}, \bar{X} + Z_{\alpha/2}\frac{S^*}{\sqrt{n}}\right]$$

where the value $Z_{\alpha/2}$ is read from the standard normal table $N(0,1)$.

Exemple 3.3.2. *The concentration of a solution of fluorescent was measured 90 times. The sample mean $\bar{X} = 4.38mg|l$ and standard deviation $S^* = 0.08mg|l$ were observed.*
- Give a confidence interval for the true concentration of the solution, at the 0.95 and 0.99 confidence level.
Ans:
Since the variance σ^2 is unknown and the size n is large, the interval is of the form,

$$\mu \in \bar{X} \pm Z_{\alpha/2}\frac{S^*}{\sqrt{n}}$$

• $1-\alpha = 0.95 \Rightarrow Z_{\alpha/2} = Z_{0.025} = 1.96$

$$\mu \in [4.38 - 1.96\frac{0.08}{\sqrt{90}}; 4.38 + 1.96\frac{0.08}{\sqrt{90}}]$$

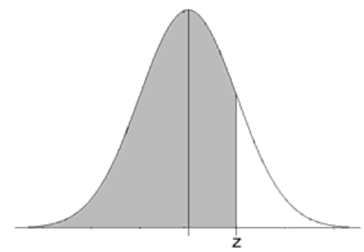
$$\mu \in [4.363; 4.397]$$

• $1-\alpha = 0.99 \Rightarrow Z_{\alpha/2} = Z_{0.005} = 2.5758$

$$\mu \in [4.38 - 2.5758\frac{0.08}{\sqrt{90}}; 4.38 + 2.5758\frac{0.08}{\sqrt{90}}]$$

$$\mu \in [4.358, 4.402].$$

Standard Normal Cumulative Probability Table



Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	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
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Confidence interval of the variance

Case 1 : the mean μ is known

The best estimator of the variance is the statistic

$$T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

The random variable $\chi^2(n) = \frac{nT}{\sigma^2}$ follows a chi-square distribution with n degrees of freedom.
A probability interval for the random variable $\chi^2(n)$ is

$$P\left(\chi^2_{1-\alpha/2,n} \leq \frac{nT}{\sigma^2} \leq \chi^2_{\alpha/2,n}\right) = 1 - \alpha$$
$$P\left(\frac{nT}{\chi^2_{\alpha/2,n}} \leq \sigma^2 \leq \frac{nT}{\chi^2_{1-\alpha/2,n}}\right) = 1 - \alpha$$

(the limits of the interval are read from the table of the chi-square distribution)

$$IC_\alpha(\sigma^2) = \left[\frac{nT}{\chi^2_{\alpha/2,n}}, \frac{nT}{\chi^2_{1-\alpha/2,n}}\right]$$

Exemple 3.3.3. *Let X be a random variable following the normal distribution $N(40, \sigma^2)$. To estimate the variance, we take a sample of size $n = 25$ and calculate the value of the statistic T (defined before) for this sample. The result is $t = 12$.*
- Give a confidence interval of σ^2 for confidence level: $1 - \alpha = 0.95$.

$$IC_{0.05}(\sigma^2) = \left[\frac{25 \times 12}{40.65}, \frac{25 \times 12}{13.12}\right] = [7.381, 22.866]$$

Case 2 : the mean μ is estimated

The unbiased estimator of the variance is the statistic

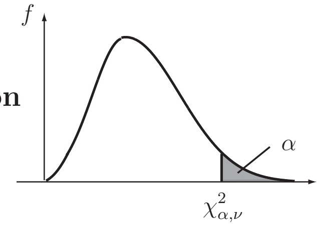
$$S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

The random variable $\chi^2_{n-1} = \frac{(n-1)S^{*2}}{\sigma^2}$ follows a chi-square distribution with $(n-1)$ degrees of freedom.

$$IC_\alpha(\sigma^2) = \left[\frac{(n-1)S^{*2}}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^{*2}}{\chi^2_{\alpha/2,n-1}}\right]$$

Here is the table of the χ^2 distribution

Percentage Points $\chi^2_{\alpha,\nu}$ of the χ^2 distribution



α	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
v											
1	0.00	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.01	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.28
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	31.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

Chapter 4

Statistical tests

Given a hypothesis H_0 concerning a population. On the basis of the results of samples taken from this population, we are led to accept or reject the hypothesis H_0. These decision rules are known as statistical tests.
H_0 denotes the null hypothesis and by H_1 we note the hypothesis called alternative hypothesis.
The final result of a statistical test is :
A case when H_0 is true and H_1 is false otherwise when H_0 is false and H_1 is true.

SECTION 4.1

Homogeneity tests

Using a sample of size n_1 and a sample of size n_2 which are independent (can be of the same population or different ones), the test makes it possible to decide :

$$\begin{cases} H_0 : & \theta_0 = \theta_1 \\ H_1 : & \theta_0 \neq \theta_1 \end{cases}$$

where θ_0 et θ_1 are the two values of the same parameter.

4.1.1 Comparison of two means (Test of Student)

Let X and Y be two normal independent random variables ($X \rightsquigarrow N(\mu_1, \sigma_1^2)$ and $Y \rightsquigarrow N(\mu_2, \sigma_2^2)$). We construct two samples of the same distribution as X ,Y : $\{X_1, X_2, ..., X_{n_1}\}$ and $\{Y_1, Y_2, ..., Y_{n_2}\}$ which are independent.

We want to decide whether the means μ_1 and μ_2 are equals or significantly different, to obtain this decision we use the test of Student.

Case one : known variance

We declare the null hypothesis with risk level α,

$$\begin{cases} H_0: & \mu_1 = \mu_2 \\ H_1: & \mu_1 \neq \mu_2 \end{cases}$$

We know that $\bar{X} \rightsquigarrow N(\mu_1, \frac{\sigma_1^2}{n_1})$ and $\bar{Y} \rightsquigarrow N(\mu_2, \frac{\sigma_2^2}{n_2})$, so the decisive variable:

$$u = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_2}}} \rightsquigarrow N(0, 1)$$

- We accept H_0 (We reject H_1), i.e., there is no significant difference between the means of two samples if:

  $$u \in [-Z_{\alpha/2}, Z_{\alpha/2}],$$

  the value $Z_{\alpha/2}$ is read from the standard normal table like we have seen before: $Z_{\alpha/2} = \phi^{-1}(1 - \frac{\alpha}{2})$.

- We reject H_0 (We accept H_1), i.e., there is a significant difference between the two means if

  $$u \notin [-Z_{\alpha/2}, Z_{\alpha/2}]$$

Exemple 4.1.1. *Let X and Y be two normal random variables with standard deviation 2.8 and 3.1 respectively, a sample $(X_1, X_2..)$ of size 120 with mean weight $48.53g$ and a sample $(Y_1, Y_2, ...)$ of size 270 of mean $50.08g$ are taken.*
At the 5% risk level (risk of error), is there a difference between the mean weights of the packets?
Ans:

Sample 1	*Sample 2*
$n_1 = 120$	$n_2 = 270$
$\bar{X} = 48.53$	$\bar{Y} = 50.08$
$\sigma_1 = 2.8$	$\sigma_2 = 3.1$

This is the test $H_0 : \mu_1 = \mu_2$

$$u = \frac{48.53 - 50.08}{\sqrt{\frac{(2.8)^2}{120} + \frac{(3.1)^2}{270}}} = -4.88$$

$$Z_{\alpha/2} = \phi^{-1}(1 - 0.025) = \phi^{-1}(0.975)$$

from the standard normal table we find $Z_{\alpha/2} = 1.96$, clearly we have $u \notin [-1.96, 1.96]$ so we reject H_0 (accept H_1), which means $\mu_1 \neq \mu_2$

case two: estimated variance

(a): n_1 and n_2 are greater than 30

In this case we replace σ_1 and σ_2 by its estimators, we find

$$u = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^{*2}}{n_1} + \frac{S_2^{*2}}{n_2}}}$$

where $S_1^{*2} = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$ and $S_1^{*2} = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$.
we accept H_0 when $u \in [-Z_{\alpha/2}, Z_{\alpha/2}]$ and reject it when $u \notin [-Z_{\alpha/2}, Z_{\alpha/2}]$.

(b): n_1 and n_2 are lesser than 30 and $\sigma_1 = \sigma_2$

Here we find the common estimator of the variance:

$$S_c^2 = \frac{(n_1 - 1)S_1^{*2} + (n_2 - 1)S_2^{*2}}{n_1 + n_2 - 2}$$

so we obtain

$$u = \frac{\bar{X} - \bar{Y}}{S_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

when u follows student distribution with $n_1 + n_2 - 2$ degree of liberty $u \rightsquigarrow T(n_1 + n_2 - 2)$, We accept H_0 if $u \in [-t_{\alpha/2,k}, t_{\alpha/2,k}]$, the value $t_{\alpha/2,k}$ is read in Student table with $k = n_1 + n_2 - 2$ degree of liberty.

(c): n_1 and n_2 are lesser than 30 and $\sigma_1 \neq \sigma_2$

In this case we estimate both variances by its estimators we obtain

$$u = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^{*2}}{n_1} + \frac{S_2^{*2}}{n_2}}} \rightsquigarrow T(v)$$

where $v = \dfrac{\frac{S_1^{*2}}{n_1} + \frac{S_2^{*2}}{n_2}}{\frac{(n_1-1)S_1^{*4}}{n_1^4} + \frac{(n_2-1)S_2^{*4}}{n_2^4}}$

We accept H_0 if $u \in [-t_{\alpha/2,v}, t_{\alpha/2,v}]$, the value $t_{\alpha/2,v}$ is read in Student table with v degree of liberty.

Exemple 4.1.2. *The weight of a medicine packaged in boxes is distributed according to a normal distribution $N(\mu, \sigma^2)$. Two samples of respective sizes $n_1 = 12$ and $n_2 = 18$ have means 22.235 g and 21.988 g and standard deviation 0.18 g and 0.23 g respectively.*
is there difference between the mean weights of the two samples for a risk level 5%?
Ans

Sample 1	*Sample 2*
$n_1 = 12$ $\bar{X} = 22.235$ $S_1^* = 0.18$	$n_2 = 18$ $\bar{Y} = 21.988$ $S_2^* = 0.23$

We have the same value of σ^2, we find the common estimator

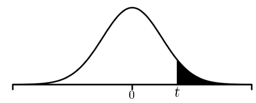
$$S_c = \sqrt{\frac{(12-1)(0.18)^2 + (18-1)(0.23)^2}{12+18-2}} = 0.21177$$

then,

$$u = \frac{22.235 + 21.988}{0.21177\sqrt{\frac{1}{12} + \frac{1}{18}}} = 3.129$$

In the Student's table, we find $t_{\alpha/2,n_1+n_2-2} = t_{0.025,28} = 2.048$
we can see that $u \notin [-2.048, 2.048]$ so we reject H_0.

Critical Values for Student's t-Distribution.



df					Upper Tail Probability: $\Pr(T > t)$					
	0.2	0.1	0.05	0.04	0.03	0.025	0.02	0.01	0.005	0.0005
1	1.376	3.078	6.314	7.916	10.579	12.706	15.895	31.821	63.657	636.619
2	1.061	1.886	2.920	3.320	3.896	4.303	4.849	6.965	9.925	31.599
3	0.978	1.638	2.353	2.605	2.951	3.182	3.482	4.541	5.841	12.924
4	0.941	1.533	2.132	2.333	2.601	2.776	2.999	3.747	4.604	8.610
5	0.920	1.476	2.015	2.191	2.422	2.571	2.757	3.365	4.032	6.869
6	0.906	1.440	1.943	2.104	2.313	2.447	2.612	3.143	3.707	5.959
7	0.896	1.415	1.895	2.046	2.241	2.365	2.517	2.998	3.499	5.408
8	0.889	1.397	1.860	2.004	2.189	2.306	2.449	2.896	3.355	5.041
9	0.883	1.383	1.833	1.973	2.150	2.262	2.398	2.821	3.250	4.781
10	0.879	1.372	1.812	1.948	2.120	2.228	2.359	2.764	3.169	4.587
11	0.876	1.363	1.796	1.928	2.096	2.201	2.328	2.718	3.106	4.437
12	0.873	1.356	1.782	1.912	2.076	2.179	2.303	2.681	3.055	4.318
13	0.870	1.350	1.771	1.899	2.060	2.160	2.282	2.650	3.012	4.221
14	0.868	1.345	1.761	1.887	2.046	2.145	2.264	2.624	2.977	4.140
15	0.866	1.341	1.753	1.878	2.034	2.131	2.249	2.602	2.947	4.073
16	0.865	1.337	1.746	1.869	2.024	2.120	2.235	2.583	2.921	4.015
17	0.863	1.333	1.740	1.862	2.015	2.110	2.224	2.567	2.898	3.965
18	0.862	1.330	1.734	1.855	2.007	2.101	2.214	2.552	2.878	3.922
19	0.861	1.328	1.729	1.850	2.000	2.093	2.205	2.539	2.861	3.883
20	0.860	1.325	1.725	1.844	1.994	2.086	2.197	2.528	2.845	3.850
21	0.859	1.323	1.721	1.840	1.988	2.080	2.189	2.518	2.831	3.819
22	0.858	1.321	1.717	1.835	1.983	2.074	2.183	2.508	2.819	3.792
23	0.858	1.319	1.714	1.832	1.978	2.069	2.177	2.500	2.807	3.768
24	0.857	1.318	1.711	1.828	1.974	2.064	2.172	2.492	2.797	3.745
25	0.856	1.316	1.708	1.825	1.970	2.060	2.167	2.485	2.787	3.725
26	0.856	1.315	1.706	1.822	1.967	2.056	2.162	2.479	2.779	3.707
27	0.855	1.314	1.703	1.819	1.963	2.052	2.158	2.473	2.771	3.690
28	0.855	1.313	1.701	1.817	1.960	2.048	2.154	2.467	2.763	3.674
29	0.854	1.311	1.699	1.814	1.957	2.045	2.150	2.462	2.756	3.659
30	0.854	1.310	1.697	1.812	1.955	2.042	2.147	2.457	2.750	3.646
31	0.853	1.309	1.696	1.810	1.952	2.040	2.144	2.453	2.744	3.633
32	0.853	1.309	1.694	1.808	1.950	2.037	2.141	2.449	2.738	3.622
33	0.853	1.308	1.692	1.806	1.948	2.035	2.138	2.445	2.733	3.611
34	0.852	1.307	1.691	1.805	1.946	2.032	2.136	2.441	2.728	3.601
35	0.852	1.306	1.690	1.803	1.944	2.030	2.133	2.438	2.724	3.591
36	0.852	1.306	1.688	1.802	1.942	2.028	2.131	2.434	2.719	3.582
37	0.851	1.305	1.687	1.800	1.940	2.026	2.129	2.431	2.715	3.574
38	0.851	1.304	1.686	1.799	1.939	2.024	2.127	2.429	2.712	3.566
39	0.851	1.304	1.685	1.798	1.937	2.023	2.125	2.426	2.708	3.558
40	0.851	1.303	1.684	1.796	1.936	2.021	2.123	2.423	2.704	3.551
41	0.850	1.303	1.683	1.795	1.934	2.020	2.121	2.421	2.701	3.544
42	0.850	1.302	1.682	1.794	1.933	2.018	2.120	2.418	2.698	3.538
43	0.850	1.302	1.681	1.793	1.932	2.017	2.118	2.416	2.695	3.532
44	0.850	1.301	1.680	1.792	1.931	2.015	2.116	2.414	2.692	3.526
45	0.850	1.301	1.679	1.791	1.929	2.014	2.115	2.412	2.690	3.520
46	0.850	1.300	1.679	1.790	1.928	2.013	2.114	2.410	2.687	3.515
47	0.849	1.300	1.678	1.789	1.927	2.012	2.112	2.408	2.685	3.510
48	0.849	1.299	1.677	1.789	1.926	2.011	2.111	2.407	2.682	3.505
49	0.849	1.299	1.677	1.788	1.925	2.010	2.110	2.405	2.680	3.500
50	0.849	1.299	1.676	1.787	1.924	2.009	2.109	2.403	2.678	3.496
60	0.848	1.296	1.671	1.781	1.917	2.000	2.099	2.390	2.660	3.460
70	0.847	1.294	1.667	1.776	1.912	1.994	2.093	2.381	2.648	3.435
80	0.846	1.292	1.664	1.773	1.908	1.990	2.088	2.374	2.639	3.416
90	0.846	1.291	1.662	1.771	1.905	1.987	2.084	2.368	2.632	3.402
100	0.845	1.290	1.660	1.769	1.902	1.984	2.081	2.364	2.626	3.390
120	0.845	1.289	1.658	1.766	1.899	1.980	2.076	2.358	2.617	3.373
140	0.844	1.288	1.656	1.763	1.896	1.977	2.073	2.353	2.611	3.361
180	0.844	1.286	1.653	1.761	1.893	1.973	2.069	2.347	2.603	3.345
200	0.843	1.286	1.653	1.760	1.892	1.972	2.067	2.345	2.601	3.340
500	0.842	1.283	1.648	1.754	1.885	1.965	2.059	2.334	2.586	3.310
1000	0.842	1.282	1.646	1.752	1.883	1.962	2.056	2.330	2.581	3.300
∞	0.842	1.282	1.645	1.751	1.881	1.960	2.054	2.326	2.576	3.291
	60%	80%	90%	92%	94%	95%	96%	98%	99%	99.9%
					Confidence Level					

Note: $t(\infty)_{\alpha/2} = Z_{\alpha/2}$ in our notation.

4.1.2 Comparison of two variances (Test of Fisher)

Let $\{X_1, X_2, ..., X_{n_1}\}$ and $\{Y_1, Y_2, ..., Y_{n_2}\}$ be two independent samples of size n_1 and n_2 derived from normal distribution $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively, the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ $(H_1 : \sigma_1^2 \neq \sigma_2^2)$.

Case one: known mean

Under H_0, the decisive variable :

$$u = \begin{cases} \dfrac{T_1}{T_2} & \text{if } T_1 > T_2 \\ \\ \dfrac{T_2}{T_1} & \text{if } T_2 > T_1 \end{cases}$$

where,

$$T_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}(X - \mu_1)^2 \quad T_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(Y - \mu_2)^2$$

and u follows Fisher Snedecor distribution of n_1 and n_2 degree of liberty ($u \rightsquigarrow F(n_1, n_2)$), which is the quotient of two variables of chi-square distribution of n_1 and n_2 degree of liberty respectively.
We accept H_0 (reject H_1) if $u < F_{\alpha,(n_1,n_2)}$ and we reject it (accept H_1) otherwise, i.e., $u > F_{\alpha,(n_1,n_2)}$
We read the value $F_{\alpha,(n_1,n_2)}$ from the Fisher Snedecor distribution table.

Case one: estimated mean

In this case we use the estimator S^{*2} instead of T, then, the decisive distribution is

$$u = \begin{cases} \dfrac{S_1^{*2}}{S_2^{*2}} & \text{if } S_1^{*2} > S_2^{*2} \\ \\ \dfrac{S_2^{*2}}{S_1^{*2}} & \text{if } S_2^{*2} > S_1^{*2} \end{cases}$$

where $u \rightsquigarrow F(n_1 - 1, n_2 - 1)$
then we accept H_0 (reject H_1) if $u < F_{\alpha,(n_1-1,n_2-1)}$ and we reject it (accept H_1) otherwise $(u > F_{\alpha,(n_1-1,n_2-1)})$

Exemple 4.1.3. *We consider two independent samples of size 11 and 9 respectively, we find the variance as 114.96 and 23.78. test $H_0 : \sigma_1^2 = \sigma_2^2$ at risk level 5%*
Ans :
we have $u \rightsquigarrow F(10,8)$,

$$u = \frac{S_1^{*2}}{S_2^{*2}} = \frac{114.96}{23.78} = 4.834$$

$F_{0.05,(10,8)} = 3.35$. *We reject H_0 since $u > F_{0.05,(10,8)}$.*

F Distribution Table

d.f.N. = degrees of freedom in numerator
d.f.D. = degrees of freedom in denominator

	α = 0.005																		
	d.f.N.																		
d.f.D.	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***12***	***15***	***20***	***24***	***30***	***40***	***60***	***120***	**∞**
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	199.5
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	41.83
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7.08
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	5.19
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	4.64
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	4.23
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	3.90
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	3.65
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	3.44
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	3.26
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	3.11
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	2.98
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	2.87
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	2.78
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	2.69
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	2.61
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	2.55
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	2.48
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	2.43
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.63	3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	2.38
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	2.33
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56	3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	2.25
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	2.29
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48	3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	2.24
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	2.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	1.93
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	1.69
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	1.43
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	1.00

F Distribution Table

	α = 0.01																		
	d.f.N.																		
d.f.D.	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99,48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

F Distribution Table

	$\alpha = 0.025$																		
	d.f.N.																		
d.f.D.	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***12***	***15***	***20***	***24***	***30***	***40***	***60***	***120***	**∞**
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.69	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

F Distribution Table

	α = 0.05																		
	d.f.N.																		
d.f.D.	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***12***	***15***	***20***	***24***	***30***	***40***	***60***	***120***	**∞**
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

F Distribution Table

d.f.D.	α = 0.10																		
	d.f.N.																		
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***12***	***15***	***20***	***24***	***30***	***40***	***60***	***120***	**∞**
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00