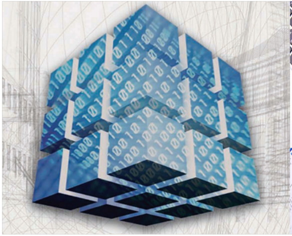
CHAPTER 3: RANDOM VARIABLES AND DISTRIBUTIONS

STATISTICS (FERM)
Lecturer: Nguyen Minh Quan, PhD
quannm@hcmiu.edu.vn




CONTENTS

Continuous Random variables

- Let X be a random variable such that its set of possible values is an interval.
- We say that X is a continuous random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)dx.$$

The function $f(x)$ is called the probability density function of the random variable X.

- Since X must assume some value, the nonnegative $f(x)$ must satisfy

$$P(X \in (-\infty, \infty)) = 1 \Rightarrow \int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Continuous Random variables

- Letting $B = [a, b]$

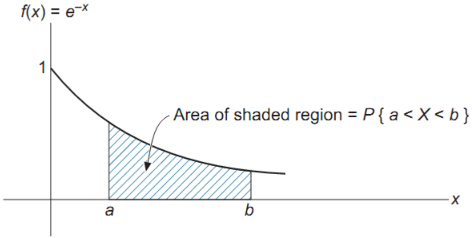
$$P(a \leqslant X \leqslant b) = \int_a^b f(x)\,dx.$$

- The probability that a continuous random variable will assume any particular value is zero (that is $P(X = a) = 0$). And

$$F(x) = P\{X \in (-\infty, x]\} = \int_{-\infty}^{x} f(t)\,dt \Rightarrow \frac{d}{dx}F(x) = f(x).$$

Continuous Random variables

Example with $f(x) = e^{-x}$.



The probability density function $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$

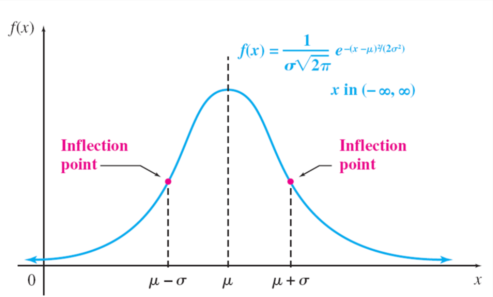
Normal Random Variables

Definition

X is a normal random variable (or simply that X is normally distributed, or X is a Gaussian random variable) with parameters μ and σ^2 if the density of X is given by

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

This density function is a bell-shaped curve that is symmetric around μ.



Normal Random Variables

Remark: if X is normally distributed with parameters μ and σ^2 then $Y = (X - \mu)/\sigma$ is normally distributed with parameters 0 and 1. Such a random variable Y is said to have the standard or unit normal distribution.

Exercise

Show that if X is normally distributed with parameters μ and σ^2 then $Y = \alpha X + \beta$ is normally distributed with parameters $\alpha\mu + \beta$ and $\alpha^2\sigma^2$.

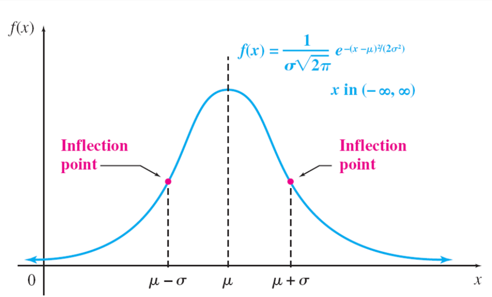
Normal Distributions

Definition

X is a normal random variable (or simply that X is normally distributed, or X is a Gaussian random variable) with parameters μ and σ^2 if the density of X is given by

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

This density function is a bell-shaped curve that is symmetric around μ.



Normal Distributions

Theorem

The expected value and variance are

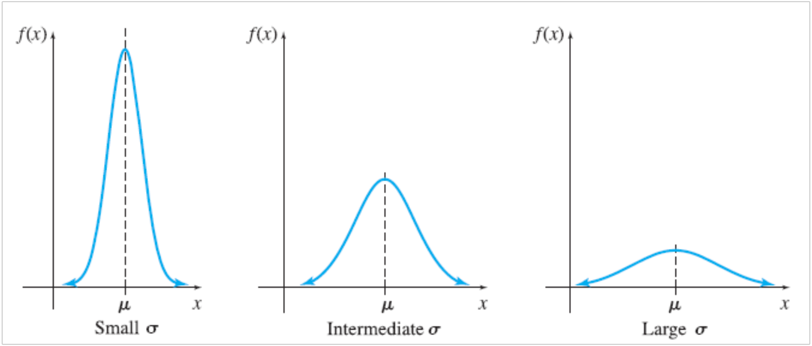
$$E(X) = \mu, Var(X) = \sigma^2.$$

Probabilities for a normal distribution is given by

$$P(a \leqslant X \leqslant b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-(x-\mu)^2/(2\sigma^2)} dx$$

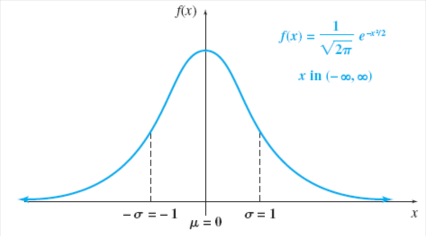
Normal Distributions

A larger value of σ produces a flatter normal curve, while smaller values of σ produce more values near the mean, resulting in a taller normal curve.



Standard Normal Distribution

The standard normal which has $\mu = 0$ and $\sigma = 1$ is called the standard normal distribution.



Standard normal distribution

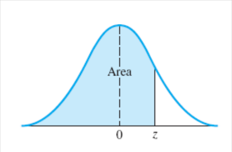
Note that

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \Phi(b) - \Phi(a)$$

where $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.

Standard Normal Distribution

$\Phi(z)$ is the area of the shaded region in the below figure:



The Table A1 gives the values of $\Phi(z)$.

Example

If X is a standard normal random variable, then

$$P(0 \leq X \leq 1.5) = \Phi(1.5) - \Phi(0) = 0.9332 - 0.5 = 0.4332$$

$$P(-0.5 \leq X \leq 1.45) = \Phi(1.45) - \Phi(-0.5) = 0.9265 - 0.3085 = 0.6180$$

Method for other normal distributions but non-standard normal?

Normal Distribution

z-Scores Theorem

Suppose a normal distribution has mean μ and standard deviation σ. The area under the associated normal curve that is to the left of the value x is exactly the same as the area to the left of

$$z = \frac{x - \mu}{\sigma}$$

for the standard normal curve.

Namely, if X is a normal random variable with mean μ and standard deviation σ then $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable.

Normal Distribution

Example

If X is a normal random variable with mean $\mu = 3$ and variance $\sigma^2 = 16$, find
(a) $P(X < 11)$.
(b) $P(X > -1)$.
(c) $P(2 < X < 7)$

Solution

$$(a) \qquad P(X < 11) = P\left(\frac{X-3}{4} < \frac{11-3}{4}\right) = P(Z < 2)$$

Therefore,

$$P(X < 11) = \Phi(2) = 0.9772$$

($Z = \dfrac{X-3}{4}$ is the standard normal random variable!)

Normal Distribution

Solution (Cont.)

$$(b) \qquad P\left(X > -1\right) = P\left(\frac{X-3}{4} > \frac{-1-3}{4}\right) = P\left(Z > -1\right)$$

Hence,

$$P\left(X > -1\right) = 1 - P\left(Z < -1\right) = 1 - 0.1587 = 0.8413$$

$$(c) \quad P\left(2 < X < 7\right) = P\left(\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right) = P\left(-1/4 < Z < 1\right)$$

$$P\left(2 < X < 7\right) = \Phi\left(1\right) - \Phi\left(-0.25\right) = 0.8413 - 0.4013 = 0.4400$$

(Note: $Z = \dfrac{X-3}{4}$ is standard normal!)

Normal Distribution

Example

The scores on an achievement test given to 100,000 students are normally distributed with mean 500 and standard deviation 100. What should the score of a student be to place him among the top 10% of all students?

Solution:

Letting X be a normal random variable with mean 500 and standard deviation 100, we must find x so that $P(X \geq x) = 0.10$ or $P(X < x) = 0.90$. This gives

$$P\left(\frac{X-500}{100} < \frac{x-500}{100}\right) = 0.90$$

Thus, $\Phi\left(\frac{x-500}{100}\right) = 0.90$.

From Table 1 of the Appendix, we have that $\Phi(1.28) \approx 0.8997$, implying that $(x-500)/100 \approx 1.28$. This gives $x \approx 628$.

Normal Distribution

Example: Investment

The annual rate of return for a share of a specific stock is a normal random variable with mean 10% and standard deviation 12 %. Ms. Lan buys 100 shares of the stock at a price of \$60 per share. What is the probability that after a year her net profit from that investment is at least \$750? Ignore transaction costs and assume that there is no annual dividend.

Solution:

Let r be the rate of return of this stock. The random variable r is normal with $\mu = 0.10$ and $\sigma = 0.12$. Let X be the price of the total shares of the stock that Ms. Lan buys this year. We are given that $X = 6000$. Let Y be the total value of the shares next year. The desired probability is

$$P(Y - X \geqslant 750) = P\left(\frac{Y - X}{X} \geqslant \frac{750}{X}\right) = P(r \geqslant 0.0125)$$

$$= P(Z \geqslant 0.21) = 1 - \Phi(0.21) = 1 - 0.5832 = 0.4168$$

Normal Distribution

Example

Suppose that a Scottish soldiers chest size is normally distributed with mean 39.8 and standard deviation 2.05 inches, respectively.
(a) Find the probability that a randomly selected Scottish soldier has a chest of 40 or more inches.
(b) What is the probability that of 20 randomly selected Scottish soldiers, five have a chest of at least 40 inches?.

Hint

$$\text{(a) } \mathrm{P}(\mathrm{X} \geq 40) = P\left(Z \geq \frac{40 - 39.8}{2.05}\right) = P(Z \geq 0.1) = 1 - \Phi(0.1) = 0.46$$

$$\text{(b)} \quad \binom{20}{5} p^5 (1-p)^{15} = \binom{20}{5} 0.46^5 (1 - 0.46)^{15} \approx 0.03$$

Normal Distribution

Theorem

If X_i, $i = 1, 2, \cdots, n$ are independent RVs with X_i being normal with mean μ_i and variance σ_i^2, then normal random variable $X = \sum_{i=1}^{n} X_i$ having mean μ and variance σ^2, where

$$\mu = \sum_{i=1}^{n} \mu_i, \sigma^2 = \sum_{i=1}^{n} \sigma_i^2$$

Proof:

Use the moment generating function of X and show that

$$\phi(t) = E\left[e^{tX}\right] = e^{\mu t + \sigma^2 t^2/2}.$$

And note that

$$\phi^{(n)}(0) = E[X^n]$$

Normal Distribution

Example

Data from the National Oceanic and Atmospheric Administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.
(a) Find the probability that the total precipitation during the next 2 years will exceed 25 inches.
(b) Find the probability that next years precipitation will exceed that of the following year by more than 3 inches.
Assume that the precipitation totals for the next 2 years are independent.

Hint:

Let X_1 and X_2 be the precipitation totals for the next 2 years.
(a) $X_1 + X_2$ is normal with mean 24.16 and variance $2(3.1)^2 = 19.22$

$$P\left(X_1 + X_2 > 25\right) \approx 0.4240$$

(b) $P\left(X_1 > X_2 + 3\right) = P\left(X_1 - X_2 > 3\right) \approx 0.2469.$

Exponential distribution

Definition

A continuous random variable whose probability density function is given, for some $\lambda > 0$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \text{ if } x \geqslant 0 \\ 0 \text{ if } x < 0 \end{cases}$$

is said to be an exponential random variable (or, more simply, is said to be exponentially distributed) with parameter λ.

The cumulative distribution function $F(x)$ of an exponential random variable is given by

$$F(x) = 1 - e^{-\lambda x}, \quad x \geqslant 0.$$

Exponential distribution

Theorem

If X be an exponential random variable, then

$$F(x) = 1 - e^{-\lambda x}, x \geqslant 0$$

$$E[X] = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2}$$

Proof instructions:

Calculate the moment generating function:

$$\phi(t) = E\left[e^{tX}\right] = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda$$

Exponential distribution

Example

Suppose that every three months, on average, an earthquake occurs in California. What is the probability that the next earthquake occurs after three but before seven months?

Solution:

Let X be the time (in months) until the next earthquake; it can be assumed that X is an exponential random variable with $\lambda = 1/3$.

$$P(3 < X < 7) = F(7) - F(3) = \left(1 - e^{-7/3}\right) - \left(1 - e^{-1}\right) \simeq 0.27$$

Exponential distribution

Exercise

At an intersection there are two accidents per day, on average. What is the probability that after the next accident there will be no accidents at all for the next two days?

Solution: Let X be the time (in days) between the next two accidents. It can be assumed that X is exponential with parameter $\lambda = 2$.

$$P(X > 2) = 1 - F(2) = e^{-4}.$$

Exponential distribution

Exercise

At an intersection there are two accidents per day, on average. What is the probability that after the next accident there will be no accidents at all for the next two days?

Hint: Let X be the time (in days) between the next two accidents. It can be assumed that X is exponential with parameter $\lambda = 2$.

Exponential distribution: Memoryless property

Memoryless property

The key property of an exponential random variable is that it is memoryless, where we say that a nonnegative random variable X is memoryless if

$$P(X > s + t | X > t) = P(X > s), \quad \forall s, t \geqslant 0$$

Theorem

Exponentially distributed random variables are memoryless.

Proof instructions: Note that $P(X > x) = e^{\lambda x}, x > 0$. Thus, $P(X > s + t) = P(X > t) P(X > s)$.

Exponential distribution: Memoryless property

Example

Suppose that a number of miles that a car can run before its battery wears out is exponentially distributed with an average value of 10,000 miles.
(a) If a person desires to take a 5,000-mile trip, what is the probability that she will be able to complete her trip without having to replace her car battery?
(b) What can be said when the distribution is not exponential?

Solution:
(a) We note that the remaining lifetime (in thousands of miles) of the battery is exponential with parameter $\lambda = 1/10$.

$$P\,(remaining\ \ lifetime > 5) = 1 - F\,(5) = e^{-5\lambda} = e^{-1/2} \approx 0.604.$$

Exponential distribution: Memoryless property

(b) If the lifetime distribution F is not exponential, then the relevant probability is

$$P\left(\textit{lifetime} > t+5 \middle| \textit{lifetime} > t\right) = \frac{1-F(t+5)}{F(t)}$$

where t is the number of miles used.

Exercise

The lifetime of a TV tube (in years) is an exponential random variable with mean 10. If Jim bought his TV set 10 years ago, what is the probability that its tube will last another 10 years?

Gamma distribution

Definition

A random variable is said to have a gamma distribution with parameters (α, λ), $\lambda > 0$, $\alpha > 0$, if its density function is given by

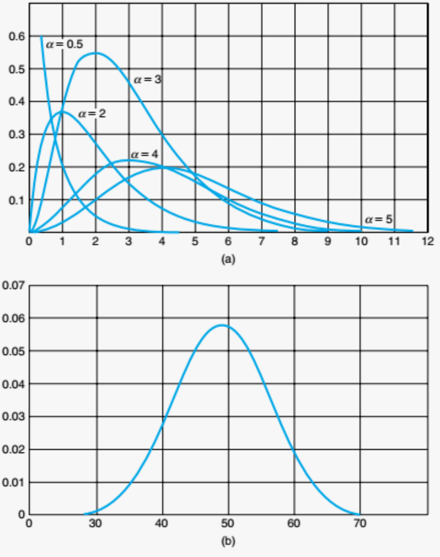
$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, x \geqslant 0 \\ 0, \quad x < 0 \end{cases}$$

where

$$\Gamma(\alpha) = \int_0^\infty \lambda e^{-\lambda x}(\lambda x)^{\alpha-1} dx$$

$$= \int_0^\infty e^{-y} y^{\alpha-1} dy \ \text{ by letting } y = \lambda x$$

Gamma distribution



Graphs of the gamma $(\alpha, 1)$ *density for (a)* $\alpha = .5, 2, 3, 4, 5$ *and (b)* $\alpha = 50$.

Gamma distribution: Properties

- When $\alpha = 1$, the gamma distribution reduces to the exponential with mean $1/\lambda$.
- $\Gamma(n) = (n-1)!$.
- If X_i, $i = 1, ..., n$ are independent gamma random variables with respective parameters (α_i, λ), then $\sum_{i=1}^{n} X_i$ is gamma with parameters $(\sum_{i=1}^{n} \alpha_i, \lambda)$.
- If X_i, $i = 1, ..., n$ are independent exponential random variables, each having rate λ, then $\sum_{i=1}^{n} X_i$ is gamma with parameters (n, λ).

Gamma distribution: Properties

Example

The lifetime of a battery is exponentially distributed with rate λ. If a stereo cassette requires one battery to operate, then the total playing time one can obtain from a total of n batteries is a gamma random variable with parameters (n, λ).

The Chi-Square Distribution

Definition

If $Z_1, Z_2, \cdots, Z_n$ are independent standard normal random variables, then X, defined by

$$X = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

is said to have a chi-square distribution with n degrees of freedom. We will use the notation

$$X \sim \chi_n^2$$

to signify that X has a chi-square distribution with n degrees of freedom.

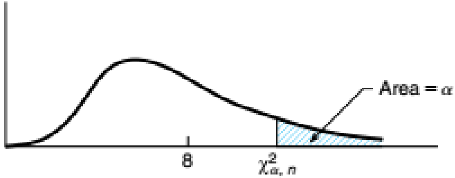
The chi-square distribution has the additive property that if X_1 and X_2 are independent chi-square random variables with n_1 and n_2 degrees of freedom, respectively, then $X_1 + X_2$ is chi-square with $n_1 + n_2$ degrees of freedom.

The Chi-Square Distribution

Definition

If X is a chi-square random variable with n degrees of freedom, then for any $\alpha \in (0, 1)$ the quantity $\chi^2_{\alpha,n}$ is defined to be such that

$$P\left(X \geqslant \chi^2_{\alpha,n}\right) = \alpha$$



The chi-square density function with 8 degrees of freedom.

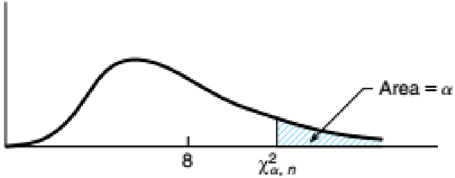
In Table $A2$ of the Appendix, we list $\chi^2_{\alpha,n}$ for a variety of values of α and n.

The Chi-Square Distribution

Definition

If X is a chi-square random variable with n degrees of freedom, then for any $\alpha \in (0, 1)$ the quantity $\chi^2_{\alpha,n}$ is defined to be such that

$$P\left(X \geqslant \chi^2_{\alpha,n}\right) = \alpha$$



The chi-square density function with 8 degrees of freedom.

In Table $A2$ of the Appendix, we list $\chi^2_{\alpha,n}$ for a variety of values of α and n.

The Chi-Square Distribution

Example

Determine $P\left(\chi^2_{26} \leqslant 12.198\right)$ when χ^2_{26} is a chi-square random variable with 26 degrees of freedom.

Solution:

Let $\alpha = P\left(\chi^2_{26} \geqslant 12.198\right)$. We have $12.198 = \chi^2_{\alpha,26}$. By Table A2, we obtain $\alpha = 0.99$. Therefore,

$$P\left(\chi^2_{26} \leqslant 12.198\right) = 1 - \alpha = 0.01.$$

The Chi-Square Distribution

Example

Find $\chi^2_{0.05,15}$

Solution:

Use Table A2, we obtain

$$\chi^2_{0.05,15} = 24.996.$$

[That is, $P\left(\chi^2_{15} \geqslant 24.996\right) = 0.05$.]

The Chi-Square Distribution

Exercise

Suppose that we are attempting to locate a target in three-dimensional space, and that the three coordinate errors (in meters) of the point chosen are independent normal random variables with mean 0 and standard deviation 2. Find the probability that the distance between the point chosen and the target exceeds 5.5911 meters.

Hint: If D is the distance, then $D = \sqrt{X_1^2 + X_2^2 + X_3^2}$, where X_i is the error in the ith coordinate.

$$P\left(D^2 > 5.911^2\right) = P\left(Z_1^2 + Z_2^2 + Z_3^2 > \frac{31.2604}{4}\right) = P\left(\chi_3^2 > 7.815\right)$$

The Relation Between Chi-Square and Gamma Random Variables

Theorem

If X is a chi-square random variable with n degrees of freedom, then X and the gamma RV with parameters $n/2$ and $1/2$ are identical.

Therefore, the density of X is given by

$$f(x) = \frac{\frac{1}{2} e^{-x/2} \left(\frac{x}{2}\right)^{(n/2)-1}}{\Gamma\left(\frac{n}{2}\right)}, \; x > 0$$

and

$$E[X] = n, Var(X) = 2n$$

The Relation Between Chi-Square and Gamma Random Variables

Example

When we attempt to locate a target in two-dimensional space, suppose that the coordinate errors are independent normal random variables with mean 0 and standard deviation 2. Find the probability that the distance between the point chosen and the target exceeds 3.

Solution:

If D is the distance and X_i, i $= 1$, 2, are the coordinate errors, then $Z_i = X_i/2$, i $= 1$, 2, are standard normal random variables, we obtain

$$P\left(D^2 > 9\right) = P\left(X_1^2 + X_2^2 > 9\right) = P\left(Z_1^2 + Z_2^2 > 9/4\right)$$

$$= P\left(\chi_2^2 > 9/4\right) = e^{-(1/2)(9/4)} = e^{-9/8} \simeq 0.3247,$$

where we used fact that $X = \chi_2^2$ is the same as the exponential distribution with parameter $1/2$.

The t-Distribution

Definition

If Z and χ_n^2 are independent random variables, with Z having a standard normal distribution and χ_n^2 having a chi-square distribution with n degrees of freedom, then the random variable T_n defined by

$$T_n = \frac{Z}{\sqrt{\chi_n^2/n}}$$

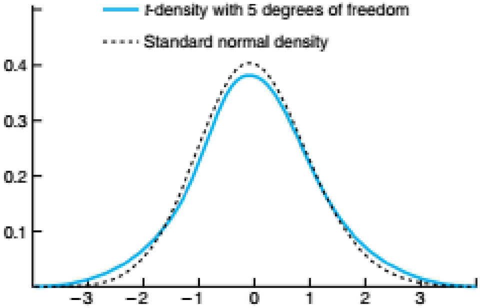
is said to have a t-distribution with n degrees of freedom.

In other words,

$$T_n \sim \frac{N(0,1)}{\sqrt{\chi_n^2/n}}$$

The t-Distribution

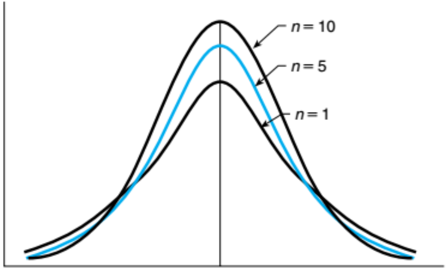
A graph of the t-density function with 5 degrees of freedom compared with the standard normal density:



Comparing standard normal density with the density of T_5.

The t-Distribution

Like the standard normal density, the t-density is symmetric about zero. In addition, as n becomes larger, it becomes more and more like a standard normal density.



Density function of T_n.

The density function of T_n for $n = 1, 5$, and 10.

The t-Distribution

Theorem

$$E\left(T_n\right) = 0, \quad n > 1$$

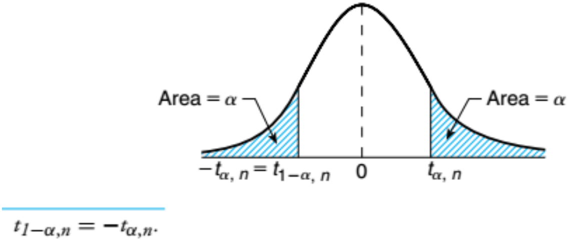
$$Var\left(T_n\right) = \frac{n}{n-2}, \quad n > 2$$

Definition of $t_{\alpha,n}$

For α, $0 < \alpha < 1$, let $t_{\alpha,n}$ be such that

$$P\left(T_n \geqslant t_{\alpha,n}\right) = \alpha$$

The t-Distribution



Remark: $-t_{\alpha,n} = t_{1-\alpha,n}$.
The values of $t_{\alpha,n}$ for a variety of values of n and α have been tabulated in Table A3 in the Appendix. For example, with $\alpha = 0.05$, $t_{0.05,5} = 2.015$, $t_{0.05,10} = 1.812$.
The t-distribution and these critical points will be used in the statistical inference methodologies.

The F-Distribution

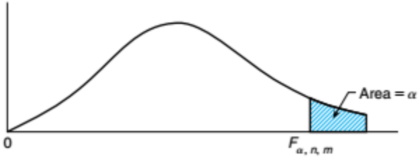
Definition

If χ_n^2 and χ_m^2 are independent chi-square random variables with n and m degrees of freedom, respectively, then the random variable $F_{n,m}$ defined by

$$F_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m}$$

is said to have an F-distribution with n and m degrees of freedom

For any $\alpha \in (0,1)$, let $F_{\alpha,n,m}$ be such that $P\left(F_{n,m} > F_{\alpha,n,m}\right) = \alpha$.



Density function of $F_{n,m}$.

The F-Distribution

The quantities $F_{\alpha,n,m}$ are tabulated in Table A4 of the Appendix for different values of n, m, and $\alpha \leq 1/2$.

Theorem

$$\frac{1}{F_{\alpha,n,m}} = F_{1-\alpha,m,n}$$

To prove that one can use

$$1 - \alpha = P\left(\frac{\chi^2_m/m}{\chi^2_n/n} \geqslant F_{1-\alpha,m,n}\right)$$

We can use the Theorem above to determine $F_{\alpha,n,m}$ for $\alpha > 1/2$. For example,

$$\frac{1}{F_{0.9,5,7}} = F_{0.1,7,5} = 1/3.37 = 0.2967$$

–END OF CHAPTER 3. THANK YOU!–