**I.4**

**REVIEW OF STATISTICAL ANALYSIS TOOLS**

**by Fernando Pérez Fontán and Perfecto Mariño Espiñeira**

**I.4.1 Introduction**

In previous chapters, we have reviewed some electromagnetic and signal theory basics, this chapter is dedicated to reviewing some statistical issues of interest. The various points touched will be illustrated, as in all chapters, by means of simulations in Matlab.

In this chapter, we are going to present some simple statistical analyses in the context of mobile communication scenarios, Part II on this book is dedicated to these scenarios. Similar analyses can also be applied to fixed systems and to effects other than shadowing and multipath (typical in mobile communications) such as those originated by the radio signal traversing the troposphere. Part III of this book is dedicated to propagation though the troposphere.

**I.4.2 Mobile communications propagation basics**

This statistical modeling review is put it in the context of terrestrial mobile communications systems. Similar techniques can also be used when dealing with tropospheric effects. Here we concentrate on the characterization of path loss, shadowing and multipath phenomena taking place in mobile communications.

Now, we would like to briefly discuss some of the characteristics of the mobile/wireless propagation channel [HerFont]. We will concentrate here on systems with base stations producing fairly large coverage areas such as in so called *macrocells*. This is the classical case, however, cell size tends to be reduced in exchange for capacity having specific propagation characteristics. The modeling techniques involved in microcell land mobile systems have also many similarities with those used in other *point-to-area coverage systems* such as sound and TV broadcasting or fixed Internet access systems.

The similarities between fixed and mobile wireless channels over the frequency bands of interest (mostly UHF bands) not only include the mechanisms giving rise to path loss. They are also subjected to shadowing and multipath effects even though these are normally much milder in the fixed systems case.

Depending on the BS/access point height, cells of larger or smaller size can be created. The classical cellular environment of tall masts above rooftops gives rise to so-called *macrocells*. As the BS antenna height becomes smaller and goes below the surrounding rooftops, so-called *microcells* are generated. BSs within buildings give rise to *picocells* (Chapter II.8). However, when satellites are used, which means much higher ‘BS antenna heights’, *megacells* are originated (Chapter II.9).

Man-made structures [Lee] such as buildings or small houses in suburban areas, with sizes ranging from a few meters to tens of meters, dramatically influence the wireless propagation channel. In urban areas, the size of structures can be even larger. Likewise, in rural and suburban environments, features such as isolated trees or groups of trees, etc. may reach similar dimensions. These features are similar or greater in size than the transmitted wavelength (*metric*, *decimetric*, *centimetric waves*) and may both block (diffraction) and scatter the radio signal causing specular and/or diffuse reflections. These contributions may reach MS by way of multiple paths, in addition to that of the direct signal. In many cases, these echoes make it possible that a sufficient amount of energy reaches the user terminal, so that the communication link is feasible. This is especially so when the direct signal is blocked. Hence, in addition to the expected distance power decay, two main effects are characteristic in mobile propagation: *shadowing* and *multipath*.

We can identify three different rates of change in the received signal as a function of the distance (spatial variations) between BS and MS, namely, *very slow variations* due to range, *slow*or*long-term variations* due to shadowing and *fast* or *short-term variations* due to multipath. Time variations, also take place. However, unless we are dealing with fixed links, they tend to be less significant than the spatial variations. We concentrate here on position dependent variations.

While in conventional macrocells BS heights are in the order of 30 m or so, and are normally set on elevated sites with no or few blocking/scattering elements in their surroundings, MS antenna heights are usually smaller than those of local, natural and man-made features. Typical values range from 1.5 or so for handheld terminals to 3 m for vehicular terminals. For other large cell radio communication systems e.g. for TV broadcasting or fixed wireless access operating in the same frequency bands, the propagation channel will present a milder behavior given that, in these cases, the receive antennas are usually directive and are normally sited well above the ground and clear of near obstacles. Both the shadowing effect on the direct signal and the amount of multipath is considerably reduced.

Other operating scenarios where both ends of the link are surrounded by obstacles are indoor communications where walls, the ceiling or the various pieces of furniture will clearly determine the propagation conditions.

Two representative and extreme scenarios may be considered:

(a) the case where a strong direct signal is available together with a number of weaker multipath echoes, i.e., *line-of-sight* (LOS) conditions; and

(b) the case where a number of weak multipath echoes is received and no direct signal is available, *non-line-of-sight* (NLOS) conditions.

*Case* (*a*) occurs in open areas or in very specific spots in city centers, in places such as crossroads or large squares with a good visibility of BS. Sometimes, there might not be a direct LOS signal but a strong specular reflection off a smooth surface such as that of a large building will give rise to similar conditions. This situation may be modeled by a Rice distribution for the variations of the received RF signal envelope: *Rice case*. Under these conditions, the received signal will be strong and with moderate fluctuations (Figure I.4.2). The Rice distribution introduced later in this chapter and is studied in more detail in Chapters II.5 and II.6. We already had a look at the complex baseband fading phenomenon associated to Rice fading in Chapter I.3.

*Case* (*b*) will typically be found in highly built-up urban environments. This is a worst-case scenario since the direct signal is completely blocked out and the overall received signal is only due to multipath, thus being weaker and subjected to marked variations (Figure I.4.1.2). This kind of situation may also occur in rural environments where the signal is obstructed by dense masses of trees: wooded areas or tree alleys. The received signal amplitude variations in this situation are normally modeled with a Rayleigh distribution: *Rayleigh case*. The Rayleigh distribution is presented later in this chapter and studied in more detail in Chapters II.5 and II.6. We had a look at the complex baseband fading phenomenon associated to Rayleigh fading in Chapter I.3.

|  |
| --- |
| rice and rayleigh |
| **Figure I.4.2** Rice and Rayleigh distributed time series. Frequency 900 MHz, mobile speed 10 m/s |

The received field strength or the received voltage may be represented in the time domain, , or in the traveled distance domain, . Figure I.4.3 shows a typical mobile communications scenario with MS driving away from BS along a radial route so that the link profile is the same as that of the route profile. The figure also shows a sketch of the received signal as a function of the distance from BS. In addition to the obvious distance dependent decay, the first thing to be noted is that the signal is subjected to strong oscillations as MS travels away from BS. Finally, not shown, are the fast variations illustrated on Figure I.4.2 which are superposed on the slow and very variations illustrated in Figure I.4.3.

|  |
| --- |
| F 1-3 modified |
| **Figure 1.3** Variations in the received signal with the movement of the mobile [2] |

For carrying out propagation channel measurements, the mobile speed, , should remain constant. Of course, there are ways around this. In such cases, the traversed distance needs to be recorded too. In our simulations, in later chapters and in the series analyzed in this chapter, we will assume a constant MS speed. For a constant terminal speed, , it is straightforward to make the conversion between the representation in the time, , and traveled distance domains, .

Variable may either be expressed in meters or in wavelengths. Based on such signal recordings plotted in the distance domain, it is possible to separate and study individually the fast and slow variations due, respectively, to multipath and shadowing, as illustrated in Figure I.4.4. Note how the solid line in Figure I.4.4, left, corresponds to the slow or local mean variations schematically illustrated in Figure I.4.3.

|  |
| --- |
| slow and fast |
| **Figure 1.4** Overall and slow variations, and fast variations after removing the slow variations [1st edition] |

Generally, the received signal variations, or , may be broken down, in a more or less artificial way, into two components [Lee],

- the slow or long-term variations: or ;and

- the fast or short-term variations: or .

Remember that the above variables represent voltages or fields, not powers or power densities. In many cases, some sort of normalization is eventually made on those variables since their actual values may be very small. In this way, reasonably large numbers are used in the calculations.

The received signal may, therefore, be described as the product of these two terms,

|  |  |
| --- | --- |
| or, alternatively, | (1.12) |

when expressed in linear units. In dB, the products become additions, i.e.,

|  |  |
| --- | --- |
| or, alternatively, | (1.13) |

With this approach, we are assuming that the fast variations are superposed on the slow variations. Figure I.3.4 illustrates a time-series where the slow variations are also plotted. The figure also shows the fast variations after removing (filtering out) the slow variations. The slow variations can be extracted from the overall variations through low-pass filtering by computing a running mean. This is equivalent to calculating the signal average for the samples within a route section of length equal to some tens of wavelengths,

|  |  |
| --- | --- |
|  | (1.14) |

Typically, lengths of to are used [Lee]. For example, for the 2 GHz()band*,* the averaging length would be. The average value, ,computed for a given route position,, is usually called the *local mean* at .

In dense urban areas, it has been observed experimentally [Lee] that the slow variations of the received signal, that is, the variations of the local mean, , follow a *lognormal distribution*(to be presented in Chapter II.6) when expressed in linear units (V, V/m, ...) or, alternatively, a *normal distribution* when expressed in logarithmic units, . We will present this case later in this chapter.

The length, , of route considered for the computation of the local mean, i.e., used to separate out the fast from the slow variations, is usually called a *small area* or *local area*. It is within a small area where the fast variations of the received signal are studied since they can be described there with well-known distributions (Rayleigh). Over longer sections of route, they need to described by means of combinations of distributions, e.g. the Suzuki distribution [Suzuki], see Chapter II.6.

Over longer route sections ranging from 50 m or 100 m to even 1 km, it that the variations of the local mean are generally studied. This extended surface is usually called a *larger area*. Typically, standard propagation models do not attempt to predict the fast signal variations. Instead they predict the mean, , the standard deviation (or *location variability*), , of the local mean variations (normal distribution in dB) and the correlation distance, , within the *larger area*.

Before low-pass filtering, the very slow variations due to the distance (also called path loss) from BS must be removed. Free-space loss, (dB), is a very common model for the range-dependent loss (Chapter I.2). The free-space loss gives rise to a distance decay in the received power following an inverse power law of exponent (Figure I.4.5), i.e.,

|  |  |
| --- | --- |
|  | (1.17) |

where indicates "proportional to". Parameter is the difference in received power expressed in dB at distances and . The above expressions show a 20 dB/decade (20 dB decrease when the distance is multiplied by 10) or 6 dB/octave (6 dB decrease when the distance is doubled) distance decay rate.

|  |
| --- |
| F 1-5 modified |
| **Figure 1.5** Received signal decay with distance: and laws |

These variations are steeper first and then become gentler for longer distances. For example, using Eq 1.16, from km 1 to km 2, a 6 dB decrease takes place. However, the same 6 dB reduction is observed from km 10 to km 20.

It has been experimentally verified that, in typical large cell mobile propagation paths, the signal's distance decay does not follow an power law (as in free space) but, rather, it presents a larger exponent. Signal decay is usually modeled by an law, i.e., being proportional to the distance risen to the power of ***.*** The values of are typically somewhere near 4, i.e., 40 dB/decade (Figure I.4.5). Widespread propagation models, discussed in Chapter II.3 such as that of Hata [4] also predict exponents close to 4.

The path-loss expressions normally provided by propagation models (Chapter II.3) are of the form

|  |  |
| --- | --- |
|  | (1.18) |

where and are dependent on the frequency and a number of other factors as listed below. Parameter is the loss at a *reference distance*, in this case, 1 km, and is the propagation decay law. Several factors, apart from the frequency and the distance that influence path loss are taken into consideration by existing propagation models affecting the expressions for and . These factors are:

- the height of the MS antenna;

- the height of the BS relative to the surrounding terrain (*effective height*);

- the *terrain irregularity* (sometimes called *undulation*, Δ*h*, or *roughness*,σt);

- the type of *land usage* (clutter) in the surroundings of MS: urban, suburban, rural, open, etc.

When calculating the path loss, all such factors must be taken into account, i.e.,

|  |  |
| --- | --- |
|  | (1.19) |

The *path loss* is defined as the loss between isotropic antennas (0 dB gain) for a given antenna separation, . Isotropic antennas do not exist in practice but are commonly used in link budget calculations since they allow the definition of the loss independently of the antennas. Then, when computing an actual link budget, the gains of the antennas gains must be introduced in the calculations.

The *path loss* is made up of three main components: a *reference loss*, typically the free-space loss, although some models like Hata’s [4] take the urban area loss as reference. Other models use the so-called *plane-earth* *loss* () as their reference (Chapter II.8).

The second component is the loss due to *terrain irregularity* (we will be discussing this in some detail in Chapter II.2) and, finally, the third component is the loss due to the *local clutter* or *local environment* where the additional loss will very much depend on the land usage in the vicinity of MS: urban, suburban, rural, open, woodland, etc.

**XXX. Common statistical distributions used in radio**

Here, we provide a brief reminder of some very basic statistical definitions. A random variable, , has a **probability density function** (pdf), , where is a non-negative, if

|  |  |
| --- | --- |
|  | (xxx) |

The associated cumulative distribution function (CDF) of , , is

|  |  |
| --- | --- |
|  | (xxx) |

that is, it gives the probability of a given value not being exceeded. Sometimes, it is more convenient to use the complementary CDF, CCDF, which gives the probability that a given value is exceeded,

|  |  |
| --- | --- |
|  | (xxx) |

Furthermore, if is continuous at , then

|  |  |
| --- | --- |
|  | (xxx) |

Intuitively, one can think of as being the probability of falling within the **infinitesimal interval** .

A pdf fulfils the following property,

|  |  |
| --- | --- |
|  | (xxx) |

Other parameters are the statistical **moments** of *,* one is the *mean* is given by

|  |  |
| --- | --- |
|  | (xxx) |

another is *mean square value*

|  |  |
| --- | --- |
|  | (xxx) |

Finally, the variance of is

|  |  |
| --- | --- |
|  | (xxx) |

where *σ* is the *standard deviation*, and the **median value**, is given by

|  |  |
| --- | --- |
|  | (xxx) |

Another parameter is the **mode** or most probable value located at the maximum of the pdf. In Figure I.4.7 a pdf and CDF are illustrated together with its moments. We now go on to review some distributions most commonly used in radio propagation studies.

**XXX The Rayleigh distribution**

Signal variations caused by multipath, in the case where the direct signal is assumed to be totally blocked, are usually represented by a Rayleigh distribution when expressed in units of voltage. In addition, if the voltage is Rayleigh distributed then the associated power follows an *exponential distribution*. The *probability density function*, pdf, of the Rayleigh distribution is given by

|  |  |
| --- | --- |
|  | (xxx) |

where is a voltage This distribution has a single parameter, its *mode* or *modal value*, . Other related parameters are given in Table I.4.1 as a function of the mode. In this chapter, we use to designate the modal value in this distribution, however, it is customary to designate it with the Greek letter which will do in the reminder of this book. However, it should not be confused with the standard deviation of this distribution.

Script **Rayleigh\_pdf\_cdf** is provided, which was used for plotting the Rayleigh pdf and CDF in Figure I.4.7 for .

|  |
| --- |
| Rayleigh parameters mode 1 |
| **Figure I.4.7** Probability density function and cumulative distribution function for a Rayleigh distribution with σ=1. Generated with **Rayleigh\_pdf\_cdf** |

By integrating the pdf, the *cumulative distribution function*, CDF, can be obtained, i.e.,

|  |  |
| --- | --- |
|  | (xxx) |

**Table I.4.1** Rayleigh distribution parameters as a function of its mode

|  |  |
| --- | --- |
| Mode |  |
| Median |  |
| Mean |  |
| RMS value |  |
| Standard deviation |  |

The CDF is very useful when computing *outage probabilities* in *link margins*. The CDF gives the probability that a given signal level is not exceeded. If this level is the system's *operation threshold*, this provides us with the probability that the signal level is equal or below such threshold, i.e., the *outage probability*. Knowing the CDF, adequate *fade margins* can also be set up.

The parameters in Table 1.1 are defined as follows:

|  |  |
| --- | --- |
|  | (xxx) |
|  | (xxx) |
|  | (xxx) |
|  | (xxx) |

where is the expectation operator.

For completeness, we show how the Rayleigh distribution can be expressed in terms of some of its parameters, other than its mode. As a function of **the mean** the pdf and CDF are as follows,

|  |  |
| --- | --- |
|  | (xxx) |

where is the *mean*.

Put now as a function of the **rms value**, the pdf and CDF have the form,

|  |  |
| --- | --- |
|  | (xxx) |

where is the *rms value*.

------

We are interested in presenting here in what context does the Rayleigh distribution normally appears in the frame of a mobile communications link. The usual case is that we are assuming a narrowband signal, here represented by a single RF frequency, that is, a continuous wave, CW. The low-pass equivalent time-varying channel frequency response, CFR, for a wireless channel subjected to multipath propagation is the result a phasor sum (see Chapter I.3), where each term corresponds to a multipath contribution, that is,

|  |  |
| --- | --- |
|  | (xxx) |

In the case of a CW or a narrowband signal, we are only interested in the time behavior of the channel at and around the carrier . In this case, the result is a time-varying complex valued time-series, , that is,

|  |  |
| --- | --- |
|  | (xxx) |

where is for example a complex voltage, and where we have the summation of multipath contributions with time-varying complex amplitudes and time-varying delays . The magnitude of term , i.e., can be modeled using a Rayleigh distribution. Normally, for convenience, this series is normalized with respect to one of its parameters: mode, rms, mean, etc. In this case, we will normalize with respect to its rms value, as it has to do with the average power.

Another commonly used distribution for this parameter is that of Rice that will be discussed later in this chapter.

Instead of working with complex base band signals, we could also work with the RF (pass-band) voltage. The relationship between a signal in the low-pass and RF domain is as follows,

|  |  |
| --- | --- |
|  | (xxx) |

which becomes

|  |  |
| --- | --- |
| VERIFICAR SI ESTO ES CIERTO | (xxx) |

The term we are interested in modeling is again the same as in the baseband case, that is, . Again, to deal with convenient magnitudes we can normalize with respect to one of Rayleigh's distribution parameters: mode, rms, mean.

**poner una ilustración *esquemática* de lo que es la portadora y la envolvente**

It is interesting to relate actual measurements or link budgets, expressed in say Watt, with our normalization values. Assume a measured time-series providing the received power when we transmitted a CW at . The instantaneous power of our RF received signal is given by

|  |  |
| --- | --- |
|  | Eq. XXX |

In this case, we have assumed an impedance . In an example below we provide an example where we use a standardized impedance value of .

We want to verify whether follows a Rayleigh distribution

|  |  |
| --- | --- |
|  | (xxx) |

We can first compute the average received power

|  |  |
| --- | --- |
|  | (xxx) |

And normalize the received voltage such that

|  |  |
| --- | --- |
|  | Eq. AAA |

Now the envelope variations of the received RF signal will show values in the vicinity of one instead of having very values of . The statistical analysis of the envelope variations, for example the outage probability or a link margin, can then be brought back to real link budget units by plugging back the normalization value used that is . This easier to do in logarithmic units.

If we equate our normalization value to a parameter of the Rayleigh distribution, for example the rms value, then value 1 of is equivalent to the distribution's rms value. The Rayleigh pdf and CDF are given above (Eq. XXX) as a function of its rms value. Moreover, if this value is equal to 1, then the pdf now becomes

|  |  |
| --- | --- |
|  | (xxx) |

and the corresponding CDF

|  |  |
| --- | --- |
|  | (xxx) |

**XXX The exponential distribution**

Associated to the Rayleigh distribution is the exponential distribution. The former represents voltages while the latter represents their associated powers. Thus, if we define a new random variable º, where is Rayleigh distributed, then is exponentially distributed whose pdf is given by

|  |  |
| --- | --- |
|  | (xxx) |

where is the mean value of the distribution. Incidentally, is also its standard deviation. In the same way as we did with the Rayleigh distribution, we can normalize p with respect to its mean value, , i.e., we define . In this case, the resulting distribution has a simple pdf given by

|  |  |
| --- | --- |
|  | (xxx) |

which is linked to the following CDF,

|  |  |
| --- | --- |
|  | (xxx) |

It is also important to note that the rms-squared parameter of the Rayleigh distribution, , is equal to the mean of the exponential distribution, .

**Example XXX Studying a Rayleigh distributed series**

File **RayleighSeries.mat** (Figure I.4.6) is supplied for analysis. It corresponds to a simulated signal (in dBm) assumed to be received under homogeneous multipath conditions. This series, even though simulated, could just as well be a measured one. One possible way of recording a measured series could be in dB units, e.g., dBm (dB relative to 1 mW) as is the case here. This could correspond to a recording with a spectrum analyzer or field strength meter. In other cases, the measured series could be given in terms of analog to digital converter (ADC) units which must be translated into voltage or power units. We are assuming here a CW transmission at 2 GHz.

File **RayleighSeries** contains a two-column matrix which includes, in the first column, the time axis in seconds and, in the second column, the power in dBm. The time axis is sampled with an interval , corresponding to a sampling frequency .

|  |
| --- |
| project11-1 |
| **Figure I.4.6**. Representation of **RayleighSeries** processed with script **fitRayleigh** |

What we will do in this example (script **fitRayleigh**) is analyze the contents of the file by computing its histogram (approximation of its pdf) and its sample CDF. Then, we will verify whether the series provided fits a Rayleigh distribution. The series is plotted in dBm in Figure I.4.6. What we want is to model the corresponding voltage. A matched load resistance, , of 50 Ω is assumed, thus, the instantaneous power can be computed by . The power and the voltage are linked through . This voltage represents the instantaneous variations in the received signal envelope. It is like an amplitude modulation of the carrier's amplitude due to multipath propagation.

The resulting values of are very small and awkward to handle. Hence, we can normalize by the voltage corresponding to the average received power as in Eq. AAA.

Here it is best to refer to the variables used in script **fitRayleigh**. First we read in our variables from the **.mat** file,

**load RayleighSeries**

**timeAxis = series11(:,1); % timeAxis in s**

**P=series11(:,2); % P in dBm**

Then we proceed to perform some unit conversions and normalize the voltage with respect to its rms value,

**p = 10.^(P/10); % now p is in mW**

**p = p/1000; % now p is in W**

**p\_mean = mean(p);**

**p\_norm = p/p\_mean; % normalize power wrt its mean**

**v = sqrt(2\*50\*p); % voltage in Volts**

**v\_rms = sqrt(mean(v.^2)); % calculate rms value**

**v\_norm = v/v\_rms; % normalize voltage wrt rms value**

Some of these intermediate results are reproduced in Figures XXX-YYY. Note the original voltages are actually very small in the order of . When normalized they become magnitudes easier to handle.

|  |  |
| --- | --- |
|  |  |
|  | Figure XX. Received signal in linear power units (W) |
|  |  |
| Figure YYY. Received signal converted to voltage (V). | Figure ZZZ. Received voltage **v\_norm** normalized with respect to its rms value |

--

Once the various magnitudes have been properly normalized, we can go on to obtain their experimental distributions and compare them with theoretical ones. One important tool to be used when computing histograms is Matlab function **hist**. This function splits the range of values of an input series into a number of separate intervals and counts the number of occurrences in each. Then another interesting function is **cumsum** which computes the accumulative sum of a series. We have used these functions in our functions below

**[CDFx,CDFy] = fCDFbins(v\_norm, 100); % compute experimental CDF**

**[CDFyTheoretical] = RayleighCDFrms\_1(CDFx); % theoretical CDF**

These functions compute the experimental CDF of series **v\_norm** and theoretical CDF for an rms value equal to 1 (Figure XXXX). To illustrate the computation of the CDF we list function **fCDFbins**. We want to split the value range into 100 bins, we proceed as follows

**[a,b]=hist(z,nBins-2);**

**a=a/length(z);**

**a=[0 a 0];**

**step=b(2)-b(1);**

**x=[b(1)-step/2 b b(length(b))+step/2];**

**y= cumsum(a);**

Note that the output of the histogram frequency is a count, not a probability. To convert it to probability we divide the count of occurrences in that particular bin by the total number of occurrences. The other output of the histogram function is a vector with the bin centers. The bin width is given by the difference between the equally spaced bin centers. Finally, we perform the cumulative sum. We should start the curve at 0 and end at 1. This function works will for most series. In the case of rain rate and attenuation series, see Part III, we usually employ the complementary CDF and there may be problems in the vicinity of value 0 which may show as a very strong delta function at the origin. We will discuss this in Part III.

The experimental and theoretical pdf are computed again using function **hist**. We have performed the steps:

**[pdfY, pdfX] = hist(v\_norm, 20);**

**pdfY = pdfY/length(v\_norm);**

**mode = 1/sqrt(2);**

**pdfTheor = RayleighHIST(pdfX,mode);**

**pdfApprox1 = 2\*pdfX.\*exp(-pdfX.^2);**

**stepp = pdfX(2) - pdfX(1);**

**pdfApprox2 = pdfApprox1\*stepp;**

First we computed the histogram and converted from counts to probabilities. Then we generated the theoretical histogram for an rms value of 1 or, equivalently, a modal value of , see Table XXX. This function integrates the area under the pdf curve within the bin value range, that is,

|  |  |
| --- | --- |
|  | (xxx) |

Moreover, if we apply the above equation to a unit rms Rayleigh distribution we have to make .

Then we plotted the pdf curve with the histogram to show that there are different things and cannot be compared directly. Finally, an approximation of the integral in Eq. XXX is performed approximately by multiplying the height of the pdf by the bin width. The obtained results are shown in Figure BBB.

|  |  |
| --- | --- |
|  |  |
| Figure CCC. Experimental and theoretical CDFs. | Figure BBB. Theoretical pdf, and histogram and modified theoretical pdf. |

-

We can also verify that the power is exponentially distributed. We can go ahead and repeat the procedure for the normalized power as follows,

**[pdfYexp, pdfXexp] = hist(p\_norm, 20);**

**pdfYexp = pdfYexp/length(p\_norm);**

**pdfTheorExp = exp(-pdfXexp); % normalized Exp pdf mean = 1**

**stepp = pdfXexp(2) - pdfXexp(1);**

**pdfTheorExp = pdfTheorExp\*stepp;**

The results are plotted in Figure XXXX.



Figure xxx. Measured and theoretical pdfs of the normalized power.

**Example xxxx. Goodness of fit test of series distribution following a Rayleigh distribution**

Script **fitRayChiTest** goes a step further from what we did in script **fitRayleigh**, we repeat most of what we already did. The difference is that before we validated the fit on a qualitative way by comparing the experimental and theoretical pdfs and CDFs. We now try to verify in an objective way whether the data actually can be fitted to a Rayleigh distribution.

|  |
| --- |
| CHI210 |
| **Figure I.4.11** Threshold in chi-square test is selected so that  [7] |

As said, from a visual comparison between the measured and the theoretical CDFs and histograms, it is clear that the agreement is quite good. Now we want to *quantify how good the fit is*. This can be achieved by means of the *chi-square goodness-of-fit test* [7] [wikipedia], see Annex i.4.XX.

From Annex I.4.xx, there are two basic elements in the chi-square test [7]. The chi-square distribution models the squared differences between a theoretical pdf and an experimental one. In Figure I.4.11 we illustrate the pdf of a chi-square distribution. We can compare the mentioned squared differences to a threshold. If this sum is small we can say the test is passed, if the sum is high, the test is failed. Details are discussed below.

First, we must define  *measure* of the difference between the values observed experimentally and the values that would be expected if the proposed pdf were correct. Second, this measure has to be compared with a *threshold*whichis determined as a function of the so-called *significance level* of the test. Usually this level is set to 1% or 5%. Note these levels correspond to the right tail of the chi-square distribution, Figure I.4.11.

Below we recall which are the steps to be followed [7] [wikipedia], for performing this test:

1.- we partition the sample space, , into the union of disjoint intervals/bins;

2.- then we compute the probability, , that an outcome falls in the -th interval under the assumption that follows the proposed distribution. Thus, if we have repetitions of the experiment, the expected number of outcomes in the -th interval is ;

3.- The chi-square measure, , is defined as the weighted difference between the observed number of outcomes, , that fall in the -th interval, and the expected number, , i.e.,

|  |  |
| --- | --- |
|  | (1.46) |

4.- If the fit is good, then will be small. The null hypothesis, i.e., that the measured data follows a given theoretical distribution, will be rejected if is too large, that is, if , where is the *threshold* for significance level .

The chi-square test is based on the fact that, for large , the random variable follows a chi-square distribution with degrees of freedom. Thus, the threshold, , can be computed by finding the point at which , (Figure I.4.11), where is a chi-square random variable with *degrees of freedom*, DoF, Annex I-4-xxx.

The thresholds for the 1% and 5% levels of significance and different degrees of freedom are given in Table I.4.2. The number of DoFs is , that is, the number of intervals or bins minus one. It is recommended that, if is the number of parameters extracted from the data (e.g., mean, standard deviation, etc.), then is better approximated by a chi-square distribution with degrees of freedom. Each estimated parameter decreases the degrees of freedom by one.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table I.4.2** Thresholds for significance levels 1% and 5%, and different degrees of freedom | | | | | |
| *K* | 5% | 1% | *K* | 5% | 1% |
| 1  2  3  4  5  6  7  8  9  10  11 | 3.84  5.99  7.81  9.49  11.07  12.59  14.07  15.51  16.92  18.31  19.68 | 6.63  9.21  11.35  13.28  15.09  16.81  18.48  20.09  21.67  23.21  24.76 | 12  13  14  15  16  17  18  19  20  25  30 | 21.03  22.36  23.69  25.00  26.30  27.59  28.87  30.14  31.41  37.65  43.77 | 26.22  27.69  29.14  30.58  32.00  33.41  34.81  36.19  37.57  44.31  50.89 |

In **fitRayChiTest** we have performed this test. First we chose the number of bins to be used in the histogram, in this case **Nbins = 10**. Histograms in this case are calculated in terms of number of counts in each bin instead of dividing by the total number of counts as we did in the previous example, Figure I.4.10. The steps followed in the script were the following,

**[HvnormY, HvnormX] = hist(v\_norm, Nbins);**

**mode = 1/sqrt(2);**

**HvnormYtheoretical = RayleighHIST(HvnormX,mode);**

**HvnormYtheoretical = HvnormYtheoretical\*length(v\_norm);**

Note how we convert the theoretical histogram, i.e., the area under the pdf within the bin (a probability) to a frequency or count by multiplying by **length(v\_norm)** which is parameter in the step`-by-step procedure above, that is the number of repetitions of the experiment.

Then we go on to calculate **D2** and the number of degrees of freedom: . From **Table I.4.2** we can read out the thresholds for 5% and 1% significance levels, those thresholds are 15.51 and 20.09, respectively.

**D2 = sum((HvnormYtheoretical - HvnormY).^2./HvnormYtheoretical)**

**df = (Nbins - 1) - 1; % reduce DOFs by 1 as we extracted rms value from data**

From the above calculations we obtain **D2 = 19.0760**. This means that the test is not passed for 5% significance and is barely passed for 1%. Using

**alpha = 1 - gammainc(0.5\*D2,0.5\*df) % significance level**

we are able to figure out the actual point in the chi-square pdf we are at. We get **alpha = 0.0145**, that is, 1.45%.

|  |
| --- |
|  |
| **Figure I.4.10** Time series and Rayleigh histograms |

The above are rather disappointing results. We know that our series has been produced with a correct simulator. We are doing something wrong. What we can do is to remove those samples which are correlated to neighboring ones. Keeping them would actually mean that we are using the same sample several times, depending on the correlation time/distance in the series. Theoretically, the samples used are required to be **independent**. We will be approximating this requirement by using **uncorrelated** samples.

Removal of correlated samples can be performed by simple decimation. A measure of the correlation time/distance can be carried out by computing the **auto-covariance** of the signal and checking the sample spacing required to go from a correlation level of 1 for null sample separations to somewhere under a value of .

Briefly we provide here the mathematical definition of the cross-covariance, implemented in Matlab's function **xcov**,

|  |  |
| --- | --- |
| , | (xxx) |

here is the mathematical expectation, \* indicates complex conjugate, and is a sample shift. The cross-covariance is computed for all possible negative and positive shifts between the two series.

To normalize the covariance, the above expression is divided by the product of variances, i.e.,

|  |  |
| --- | --- |
|  | (xxx) |

which we will call **correlation coefficient** and is provided by Matlab by adding '**coeff**' to the list of inputs to **xcov**. In the case both series are the same, we get the auto-covariance. To cover all possible shifts, the result is a series of values. The normalized auto-covariance will have a maximum of one for zero shift, right in the center of the resulting series.

After computing the auto-covariance of **RayleighSeries**, we can see (Figure I.4.10) how, at a time spacing of 0.25 s, the correlation coefficient goes below . A time spacing of is equivalent to 5 samples, i.e., 5×0.05 = 0.25. In Figure I.4.10 we only show the central part of the resulting covariance function, for spacings smaller than 5 s.

|  |
| --- |
|  |
| Figure I.4.10. Normalized autocorrelation function of original time series. |

Thus, we decimate the series by a factor of 5. This can be achieved by taking one in every five samples in the series, i.e.,

**v\_norm\_dec = v\_norm(1:5:length(v\_norm))**

To prove that we have not lost any information, we plot again the CDFs of both the original and the decimated series in Figure I.4.11. We have used a double logarithmic scale plot (**loglog**) to make the differences more apparent, especially at the lower tail of the distribution. From the figure, it is clear that the distribution has not changed after decimation.

|  |
| --- |
|  |
| Figure I.4.11. CDFs of original and decimated series in double-logarithmic scale. |

Thus, we proceed to perform the chi-square test on the decimated series. We repeat the test on the decimated series (Figure I.4.12) and we get a value for equal to 9.8488, which means we passed the test for both significance levels by a very large margin. The value corresponding to the we got is **alpha = 0.2830**.

|  |
| --- |
|  |
| Figure I.4.12. Time series and mean=1 Rayleigh histograms. Decimated series. |

**XXX The Gaussian/normal distribution**

This distribution describes continuous variables with positive or negative values [ITU]. When many random effects are added together they usually give rise to a Gaussian distribution. We will see this distribution, for example, in the variations in the received signal in mobile communications due to shadowing effects or, in tropospheric propagation describing the fluctuations of a quantity around its mean value (scintillation). In addition, it is used to describe the thermal noise in any communications system. Thus will also need it when we discuss the noise effects on measurements or when computing link budgets.

It is also interesting in the modeling of some magnitudes in logarithmic units. If the log magnitude of some random variable is Gaussian distributed, its linear magnitude is lognormally distributed. The lognormal distribution will be presented in Chapter II.6. Here we remind the reader of some basic facts about the normal or Gaussian distribution that we will need in the next example.

The Gaussian pdf is given by

|  |  |
| --- | --- |
|  | (1.49) |

and the cumulative distribution is

|  |  |
| --- | --- |
|  | (1.50) |

where is the function describing the right tail of the Gaussian distribution.

It is helpful to normalize the Gaussian random variable using its mean, , and standard deviation, . The new normalized Gaussian random variable, , is defined as whose mean is zero and its standard deviation equal to one.

The Gaussian tail function is given by

|  |  |
| --- | --- |
|  | (1.51) |

Function provides an easy way of calculating the probability that random variable fulfills that . This is equivalent to calculating the area under the tail of the pdf (Figure I.4.17). Function is easily related to the error function or its complementary also available in MATLAB® (functions **qfunc**, **erf** and **erfc**) through

|  |  |
| --- | --- |
|  | (xxx) |

Table 1.4 provides some practical values of the normalized Gaussian complementary CDF ().

**Table 1.4** Relevant values of

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 0 | 0.5 | 1.282 | 10−1 |
| 1 | 0.1587 | 2.326 | 10−2 |
| 2 | 0.02275 | 3.090 | 10−3 |
| 3 | 1.350 × 10−3 | 3.719 | 10−4 |
| 4 | 3.167 × 10−5 | 4.265 | 10−5 |
| 5 | 2.867 × 10−7 | 4.753 | 10−6 |
| 6 | 9.866 × 10−10 | 5.199 | 10−7 |
|  |  | 5.612 | 10−8 |

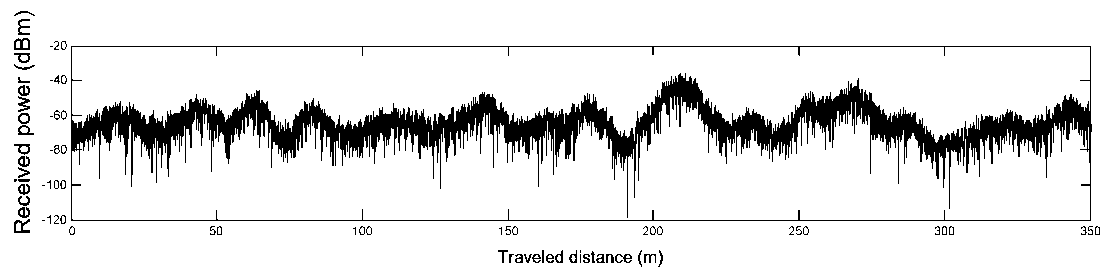
|  |
| --- |
|  |
| **Figure I.4.17** Function , area under the tail of the Gaussian distribution |

**Example xxxx Analyzing a time-series with shadowing and multipath**

The following example is implemented in script **fitSuzuki**. The name comes from the fact that we are going to analyze a time-series where there are fast fading effects (Rayleigh distributed) superposed on slow fading (Gaussian distributed). This mixed distribution is called the Suzuki distribution [Suzuki], which we will discuss in some detail in Chapter II.6.

File **SuzukiSeries.mat** is provided which includes a simulated continuous wave (CW) signal measurement with carrier frequency, . This file could be representative of a sector of a circular route at a constant distance from the transmitter so that the **nominal distance-dependent received power**,, remains constant.

File **SuzukiSeries** contains a two-column matrix where the first column represents the traveled distance in meters and the second, the received signal in dBm. The sampling spacing is . The wavelength in this case is . Note that here, the signal is sampled in the traveled distance domain.



**Figure XXX.** Plot of **SuzukiSeries** where the abscissa is in length units and the ordinate shows the received power in dBm.

From a visual observation, it is clear that both, slow and fast variations, are present. **Separation** of these two effects must be performed by computing the running mean. This is equivalent to filtering the series with a rectangular window that is slid through the series.

Matlab function **conv** (convolution) can be used for this purpose. This process gives unreliable samples at the beginning and ending of the filtered series, which have to be discarded. We know that the convolution, **C=conv(A,B)**, produces a resulting vector of length

**MAX([LENGTH(A)+LENGTH(B)-1,LENGTH(A),LENGTH(B)])**.

If we use parameter **SHAPE** in the convolution function, i.e., **C=conv(A,B,SHAPE)**, where the value of **SHAPE** is **'same'**, the function returns the central part of the convolution that is of the same size as **A**. This solves our problem and we no longer need to clip off the unreliable samples at the beginning and ending of the convolution. Another alternative would be using function **filter**. We will be using this function elsewhere on this book.

The impulse response of the running mean or sliding window, , used here as follows,

|  |  |
| --- | --- |
| , |  |

that is, a collection of Dirac deltas with ampitude, where is the window length. Note that this is a FIR filter, Chapter I.3.

Script **SuzukiSeries** loads the file and assigns its contents to our variables

**load SuzukiSeries**

**d = series12(:,1);**

**P = series12(:,2);**

The original series, **P**, is in dBm and should be converted into linear power, **p**, units (W), that is, **p = 10.^(P/10)/1000**.

A **window** size of 10λ is proposed for separating the fast and slow variations. Since we are sampling with a spacing of , the window length is . As , the window length is 0.15/4\*40=1.5 m. Other window sizes can be tested by the reader. The decision is difficult to make since, depending on the length, either multipath effects will leak into the extracted shadowing effects or the other way around. This should be studied on a case by case basis.

**wlFraction = 4; % samling fraction of wavelength**

**windowWavelengths = 10; % No. of wavelengths in running mean filte**

**windowLength = wlFraction\*windowWavelengths; % samples in window**

**W = ones(windowLength,1); % Create running mean window**

**W = W/windowLength; % Normalize window**

~~- Finally, we should model separately the slow and fast variations obtained via filtering. Models normally assume that shadowing effects are Gaussian distributed when expressed in logarithmic units. When we use linear units for voltage or power, then we end up with a log-normal distribution which is more difficult to study. We leave mixing Rayleigh and log-normal distributions directly to Chapter I.6.~~

To separate the fast and slow variations, we need to convert to linear units and apply the running mean discussed earlier. Assuming that the received power series in linear units results from the product of two components, we get

|  |  |
| --- | --- |
|  | (xxx) |

These two series vary at different rates as will be illustrated below.

The first series represents the local mean power around route point , while the second represents the superposed fast variations with normalized unit power.

In parallel, the series in dBm, can be represented as

|  |  |
| --- | --- |
|  | (xxx) |

The procedure followed to is to calculate the local mean of ,, through low pass filtering, this is carry out using

**pfilt = conv(p,W,'same');**

this variable should contain the variations due exclusively to shadowing in linear units.

Next, we can convert the resulting series to logarithmic units, ,

**Pfilt = 10\*log10(pfilt) + 30; % we now go back to dBm**

Our objective is proving that series (**Pfilt**), follows a Gaussian distribution. To do this, we carry on in much the same way as for the Rayleigh case. We first extract from the data the parameters we need to compute the theoretical distribution,

**MM = mean(Pfilt);**

**SS = std(Pfilt);**

Then we compute the histogram for the experimental data

**N\_bins = 20;**

**[a,b] = hist(Pfilt,N\_bins);**

**histBin = b(2)-b(1);**

**a = a/length(Pfilt);**

**aa = cumsum(a);**

and the theoretical histogram for comparison on the same plot. We will not carry out the goodness-of-fit test here,

**Paxis = (min(Pfilt):max(Pfilt));**

**pdf = 1/(sqrt(2\*pi)\*SS)\*exp(-0.5\*((Paxis-MM)/SS).^2);**

**fhist = pdf\*histBin;**

To compute the theoretical distribution, we can use

**F = 1-qfunc((Paxis-MM)/SS);**

In order to separate out the fast variations, we could implement a high-pass filter. Alternatively, we can operate in logarithmic units and subtract from the overall (instantaneous) series the newly extracted local mean value, i.e.,

|  |  |
| --- | --- |
|  | (xxx) |

and, finally, convert the result back to linear, that is,

**P0 = P - Pfilt;**

**vnorm = 10.^(P0/20);**

Variable **vnorm** should contain the isolated fast variations, that is superposed on a constant level.

Next we illustrate and comment the obtained results. Figure 1 shows the first 100 m of original series in dBm, Figure 2 illustrates the running mean used as averaging filter. Figure 3 depicts the instantaneous powers and the local means in linear units. Figure 4 shows the same information using dBm units. Figure 5 shows the experimental pdf (histogram) and the theoretical pdf corresponding to a mean value **MM = -62.23** (dBm) and a standard deviation **SS = 6.56** (dBm). Figure 6 shows the experimental CDF in histogram form and the theoretical Gaussian CDF. Finally, Figure 7 shows the normalized fast variations alone after being separated out from the overall variations expressed in dB. If converted to linear we would find it is exponential distributed with mean value one if converted to power or a Rayleigh distribution with rms value one if converted to voltage.

|  |  |
| --- | --- |
|  |  |
| Figure 1. Original series in dBm. First 100 m of route | Figure 2. Window used to perform running mean filtering |

-

|  |  |
| --- | --- |
|  |  |
| Figure 3. Instantaneous power series and local means in linear units. Zoom on first 100 m of MS route | Figure 4. Instantaneous power series and local means in dBm. Zoom on first 50 m of MS route |

-

|  |  |
| --- | --- |
|  |  |
| Figura 5. Measured pdf (histogram) and theoretical Gaussian pdf. Mean **MM = -62.23** (dBm) and standard deviation **SS = 6.56** (dBm). | Figura 6. Measured CDF (histogram) and theoretical Gaaussian CDF. Mean **MM = -62.23** (dBm) and standard deviation **SS = 6.56** (dBm). |

-

|  |  |
| --- | --- |
|  |  |
| Figura 7. Variaciones rápidas separadas de las variaciones totales. El valor de la media de la potencia normalizada en unidades lineales es 1. Zoom de los primeros 50 m. |  |

**XXXXX. The Rice distribution**

....................................................................

.............................................................................................

markum function .....

The Rayleigh model for signal fading is really a worst-case scenario. It postulates that the direct signal is totally blocked. Fading conditions, when the direct signal is unblocked or slightly attenuated (shadowed), can be quite well described by using the Rice distribution. This distribution is also used in tropospheric modeling since it fits the behavior of scintillation in clear air conditions, see Chapter III.xxxx.

In **fitRiceGreenstein**, we make the same assumptions as in **fitRayleigh**, the only change is that a new contribution, i.e., the direct ray between BS and MS is present. Its amplitude, , can take on values ranging from one, meaning unblocked, LOS conditions, to zero, meaning a totally blocked direct component, which is equivalent to the Rayleigh case. The Rice pdf is given by equation

|  |  |
| --- | --- |
|  | (5.16) |

Again, represents a voltage. is the modified Bessel function of first kind and zero order, which is available in Matlab-(**besseli**). This function can be calculated using the expression

|  |  |
| --- | --- |
|  |  |

Unlike in the case of a Rayleigh distribution were we changed the usual naming of variables and we replaced with (in order not to be mistaken with the standard deviation of the Rayleigh). Here we stick to the common way of naming variables and use . Thus, the reader is reminded that this parameter is not the standard deviation of the Rice distribution. Parameter is very important in both Rayleigh and Rice distributions as we shall see below.

The carrier-to-multipath ratio in dB is given by

|  |  |
| --- | --- |
|  |  |

where is the amplitude of the direct ray. In some cases, for example in satellite maritime communications, it may represent the coherent combination (complex sum) of the direct and the specular ray. Parameter is related to with the diffuse multipath power. Parameter also known as the *carrier-to-multipath ratio* (C/M). Note in Figure I.4.17 the evolution of the Rice pdf toward a Rayleigh distribution as the direct signal decreases.

|  |
| --- |
| RICE DISTRIBUTIONS |
| **Figure I.4.17** Evolution of the Rice distribution pdf toward a Rayleigh distribution as the direct signal decreases |

It is important to know that both Rayleigh and Rice distributions are related. We can see in Figure 5.17 how, as de direct signal gets smaller and smaller, the overall distribution tends to a Rayleigh. Finally, when a=0, then it actually becomes a Rayleigh distribution, that is,

|  |  |
| --- | --- |
|  | (xxx) |

we have replaced for .

Figure I.4.18 shows the time-series in file **RiceSeries.mat** we are going to analyze with script **fitRiceGreenstein**. The results after running this script are shown in Figures 5.19 and 5.20. They correspond to a Rice distribution with a value of approximately 20 dB, where the direct signal amplitude is one. Note the drastic change in received signal levels. Now the oscillations are about the zero dB level, while before they were close to –20 dB. The dynamic range of the oscillations has also changed, now the fades are less deep than in the Rayleigh case, showing how, when the direct signal is present (for example at a crossroads or an open are with direct view of BS), the channel is much milder, posing less problems for the transmitted signal. Note in Figure I.4.19 we have produced an already normalized time-series (see y-axis, it says signal relative to LOS) with parameters and which equivalent to having and . Our estimator, explained below, yields the following results **a\_estim = 0.9999**,

**sigma\_estim = 0.0708**, **k\_estim = 99.7980** and **K = 19.9912**, as expected.



Figures 5.19



Figure 5.20

Now, we discuss the procedure implemented in **findRiceParMoM** called by **fitRiceGreenstein**. Following [Gree99]. The model assumes the sum of two complex phasors,

|  |  |
| --- | --- |
|  |  |

the resulting envelope, , is Rice distributed. Locally, is assumed constant.

The **method of moments**, MoM, can be used for extracting the parameters of a Rice distribution as discussed in [Gree99]. (We have slightly change some of the parameter conventions to agree with the ones followed throughout this book, mostly that we usually will work assuming we are the baseband domain.)

First, the voltage envelope is squared, i.e., converted to power,

|  |  |
| --- | --- |
|  | (xxx) |

and two moments of the series are computed, the first one is **the mean**,

|  |  |
| --- | --- |
|  | (xxx) |

and the second moment is the **rms variations about the mean**, that is,

|  |  |
| --- | --- |
|  | (xxx) |

From these two parameters we can obtain

|  |  |
| --- | --- |
|  | (xxx) |

we equate

|  |  |
| --- | --- |
|  | (xxx) |

Finally,

|  |  |
| --- | --- |
|  | (xxx) |

which is the so call Rice factor or carrier-to-multipath ratio. This procedure is implemented in the follwoign function,

**[a\_estim, sigma\_estim, k\_estim] = findRiceParMoM(v)**

-------

**Example XXXX Rice series in power units**

Some difficulties may arise when we are given a time series in actual power levels, unlike in the previous example. So, now we have generated a series **RiceSeries2** where we have used the following parameters: and . This parameter combination corresponds to . Our simulator (Chapter II.2) references the output such that amplitude 1 means direct line-of-sight (LOS) conditions. Then we have converted the series to dB and added . This means that that level in dBm represents the LOS conditions. The series is shown in Figure XXXX.



Figure XXXX.

We find a perfect visual match between the experimental and theoretical distributions (not plotted, please run script **fitRiceGreenstein2**). In this case, we have normalized the corresponding voltage with respect to its rms value as follows

**R = 50; % Ohm, impedance**

**PrdBW = Pr - 1000; % from dBm to dBW**

**pr = 10.^(PrdBW/10); % signal in Watt**

**v = sqrt(2\*pr\*R); % signal in volts**

**vrms = sqrt(mean(v.^2));**

**v\_norm = v/vrms;**

The results obtained using the same methodology as before were a\_estim = 0.9280, sigma\_estim = 0.2635, k\_estim = 6.2019 and K = 7.9252. We see that the K factor is perfectly estimated. However, there is a difference between our estimated and original parameters.

The difference may have arisen from the normalization we used in fact the rms value of the normalized series **v\_norm**, is equal to one while its formal expression is , which is the case with our extracted parameters. There is an offset in our estimations which is mostly due to our simulator's reference level, i.e., the LOS level. When performing measurements, it is difficult to establish the LOS level, and thus, when we want to perform the normalization we have to go for an easily measurable parameter such the rms value in our series.

We have gone ahead and generated another Rice distributed time-series, see Figure XXXX. The parameters utilized in this case by our simulator are as follows: and . This parameter combination corresponds to . Then we have converted the series to dB and added . Note the increase in the series' dynamic range as the K factor decreases. Again level corresponds to LOS conditions.



Figure XXXX.

Again visual inspection of the experimental and theoretical CDFs signifies that there is a reasonable fit. However, the results obtained using the same methodology as before were a\_estim = 0.5783, sigma\_estim = 0.5769, k\_estim = 0.5026 and K = -2.9878. We see that the K factor is perfectly estimated. However, there is a difference between our estimated and original parameters. Again we have that

**sqrt(a\_estim^2+2\*sigma\_estim^2) = 1**

Again, unless we have a way of figuring out the direct signal's (LOS level) level, we will have to limit ourselves to the parameter set we have obtained with this procedure.

-------

**Example xxxx Normalized Rice series and noise**

Now we study the case where the time-series is corrupted by noise, which is the case when we analyze a measured series. We go back to our chosen normalization of time series with respect to LOS conditions, meaning that amplitude 1 means an unblocked LOS link.

The example is implemented in **fitRiceGreensteinSNR**. Again the input parameters are supplied at the start. For reference the ideal case (no noise) is also launched. We set the following parameters

**a = 1, sigma = 0.1000, noSamples = 100000** and **SNR = 10**

The Rice series corrupted by noise is produced by function **genRiceUncorrSeriesSNR**. We generate the noise series both in the in-phase and quadrature axes. The standard deviation multiplying the random Gaussian generators was computed as

**sigma\_noise = sqrt((a^2+2\*sigma^2)/(SNR \* 2));**

which corresponds to the expression

|  |  |
| --- | --- |
|  | (xxx) |

where, working in the **complex base band**, the signal power is the sum of the powers of the direct signal plus the multipath, that is , and the noise power, .

After applying the MoM estimator, we get **aSNR\_estim = 1.0098** and **sigmaSNR\_estim = 0.2445**, which pretty much coincides with **sigma\_noise = 0.2258**, calculated within **genRiceUncorrSeriesSNR**. In this case, the obtained **sigmaSNR\_estim** is totally blurred over by the noise. Figure XXX.a illustrates the CDF of an ideal Rice distributed series and those from the Rice series corrupted by noise and an ideal Rice series generated using the parameters off the estimation process. The reader is reminded that we have not carried out a goodness-of-fit test to make sure that we are actually working with Rice distributed series.

It is clear that if we want to perform a statistical analysis of fading, we will need to allow for higher SNR values if we want to perform accurate analyses. For example, if we are carrying out a measurement campaign we should either raise the transmit power or, alternatively, work at a shorter distance from the transmitter.

We see that adding noise we get another Rice distributed Rice series. Physically, the signal oscillations due to noise would be totally random while those due to multipath would be slower in a properly sampled series (correlation time, to be discussed in Chapter II.7). However, as said earlier, a fast sampled series has to be decimated all the way down so as to get uncorrelated samples. Thus, the mentioned the two different rate of change are no longer visible.

We have worked out another example with input parameters **a = 0.5; sigma = 0.3; noSamples = 100000; SNR = 10;** and the estimated parameters are **aSNR\_estim = 0.5010** and **sigmaSNR\_estim = 0.3333**, and where **generateRiceUncorrSeriesSNR**'s internal parameter **sigma\_noise = 0.1466**.

In this case the estimated value of **sigmaSNR\_estim** is not far for the actual ideal case value. The obtained CDFs are now showed in Figure xxxx.b where se observe how the

|  |  |
| --- | --- |
|  |  |
| Figure xxxx |  |

------

**xxxx. Modeling the path loss**

We have discussed path loss earlier in this chapter. The best way of presetting this loss is by means on an example.

**Example xxxx. Linear regression for extracting the path loss**

In this example, we analyze a set of data simulating a measurement campaign, in fact the data has been simulated. From the analysis we develop an **empirical model** for path loss using linear regression.

The received signal in a mobile-base link contains three types of signal variations: fast, characterized by the instantaneous signal, , and slow, characterized by the local mean, . Moreover, there exist much slower variations due to the Tx-Rx distance dependence, the path loss, which can be characterized by a larger area average or nominal power, , as illustrated in **Figure I.4.1** [Sklar].

Normally, the nominal signal power will decay according to the inverse of the distance risen to the power , which normally is close to , that is, using powers in linear units,

|  |  |
| --- | --- |
|  | (xxx) |

Remember that for the free space case.

To study the nominal power, , at a distance, the received signal must be low-pass filtered to remove the instantaneous variations, , and further filtered to remove the slow variations, , due to shadowing effects.

The above expression can also be put in logarithmic units such that

|  |  |
| --- | --- |
|  | (xxx) |

The path loss or nominal loss for a distance is obtained from the above expression if we know the magnitudes of the various elements in the link budget, i.e.,

|  |  |
| --- | --- |
|  |  |

The nominal power for a distance can be calculated using the expression

|  |  |
| --- | --- |
|  |  |

where is the Equivalent Isotropic Radiated Power, defined as the sum of the transmitter's power, antenna gain and losses (cables, etc.), and is the received antena gain.

|  |
| --- |
|  |
| **Figure 1.** Slow and fast variations and associated margins [Sklar] |

The objective of this example to process a set of pseudo-experimental data and develop an empirical propagation model like the one just discussed. The type of model extracted is very similar to the model due to Okumura [Okumura] which was later processed to produce path loss expressions such the one in Eq. XXX. by Hata [Hata]. Further information on common propagation models can be found in Chapter II.XXXX.

We are provided with a large number of value pairs, the **distance**, , versus the **received power,** , for a base-terminal mobile link. We assume that the fast signal variations have been filtered out. This meanms that only distance and shadowing effects are present.

We will try and develop a propagation model like the one described above where the averaged received power is due to the path loss. Using logarithmic units, we have

|  |  |
| --- | --- |
|  | (xxx) |

where is the power at a reference distance, in this case 1 km. If we are going to work in km, then is more convenient to represent the power at 1 km. This is the case when we work with macrocells. If the distances we are interested in are shorter, for example in meters as in microcells or picocells, then is better expressed as the power at 1 m distance. If is given in dBm, then will be put in dBm.

Superposed on the nominal received power, we will have the slow variations (in logarithmic units) due to shadowing which are modeled with a Gaussian distribution with a mean given by the previous equation and a standard deviation, (dBm). The values in the file provided (**pathloss.mat**) thus represent power, (dBm), resulting from the sum of the nominal power and the slow, shadowing variations, that is,

|  |  |
| --- | --- |
|  | (xxx) |

where is a Gaussian random variable with zero mean and standard deviation

File **pathloss** contains the following two vectors: **distkm** and **Pr**, containing the distance in km and the corresponding local average received power, (dBm).

The data in the file have been simulated (script **genPathLoss**), but we will assume that they correspond to actual measured data out of a measurement campaign.

The first step in the analysis is **ordering** the samples as a function of the distance using

**[distkm,I] = sort(distkm)**

and then reorganize **Pr** so that their correspondence is maintained using index **I**.

Figure 2 shows the values of **Pr** as a function of the distance, **distkm**. It is evident that a straight line cannot be fitted to the data. We need to make a transformation, in this case, we take the logarithm of the distance, more specifically, , as specified by the wanted model. After taking logarithms, we get the scatter plot in Figure 3. where a clear linear trend is observed.

Thus, the regression study will be performed for **Pr** in dBm and . In Figure 3 the obtained linear model is also shown toether witrh the data cloud. The fitted linear model has been obtained using Matlab functions **polyfit** and **polyval**. Further information on simple data fitting techniques can be found in Annex I.4.2.

|  |  |
| --- | --- |
|  |  |
| Figure 2. S**catter plot** of the original data in file **data13.mat**. | Figure 3. Data cloud where the abscissa has been converted to linear units and fitted linear model. |

We have asked **polyfit** to provide a first order polynomial, a straight line. The obtained results are as follows,

**P = -3.** **7852 -58.709**,

that is, the sought model is

|  |  |
| --- | --- |
|  | (xxx) |

The first parameter is the power at the reference distance, in this case, 1 km while the slope provides the decay/increase rate.

It is possible to verify how good the model is by computing the determination coefficient, Annex I.4.2. We see that it is very high,

**Model's coefficient of determination : 0.87332**

indicating the model derived by linear regression explains 87% of the data's original uncertainty.

**Figure 4** shows the **residuals**, that is, the difference between the measurements and the model. We can see how their mean is zero and their standard deviation, **sigmaL**, is .

|  |  |
| --- | --- |
|  |  |
| Figure 4. Residuals: measurements minus model. | |

It is important to verify that the residuals' distribution is Gaussian distributed. In this case, we can plot the experimental and theoretical CDFs together and observe that they are very close. To stress their similarity at least in the lower values, distribution tail, we have put the ordinates in logarithmic scale, Figure 5.

It is important to point out that we are assuming that the errors, i.e., the residuals, can actually be attributed to shadowing effects.

Traditionally, before widespread computer availability, a very convenient way of verifying the Gaussian character of a distribution was to use a special gridded paper with its axes modified in such a way that the CDF follows a straight in case the data were Gaussian distributed. If Matlab's **Statistical Toolbox** is available, the reader can perform this test using function **normplot**, **Figure 6**).

|  |  |
| --- | --- |
|  |  |
| Figure 5. measured and theoretical CDFs of the slow variations/residuals. | Figure 6. Residuals/slow variations CDF plotted on Gaussian paper. |

We now comment on some of the Matlab functions used. Matlab has a built-in function **polyfit** that fits a least-squares n-th order polynomial to data with the following syntax,

**p = polyfit(x, y, n)**

where **x** is the independent data, **y** is dependent data, **n** is order of the polynomial we want to fit to the data. Finally, **p** contains coefficients of the obtained polynomial, with the generic form

|  |  |
| --- | --- |
|  | (xxx) |

MATLAB’s **polyval** command can be used to compute a value using the obtained coefficients,

***y* = polyval(*p*, *x*)**

The pseudo-experimental data was produced using script **genPathLoss**. It uses

**y = randn(M,N)**

to generate normally distributed numbers. It returns an M-by-N matrix containing pseudorandom values drawn from the **standard** normal distribution: zero mean and unit standard deviation. We have earlier discussed the normalized Gaussian distribution.

To generate random data corresponding to other Gaussian distributions, mean **MM** and standard deviation **SS**, we further transform **y** as follows,

**z = y\*SS + M**

**1.5 Summary**

...............................................

Then, we presented very simple time-series analysis techniques which cover the basics previously introduced, these include the fast variations due to multipath and the combined effects of shadowing and multipath. Finally, we have looked into the complex envelope where we plotted its magnitude and phase. We have also presented the Rayleigh and Rice cases which correspond to harsh and benign propagation conditions.

We will now go on to learn more about the various phenomena presented through simulation. This will allow us to become acquainted with a number of fairly complicated concepts in an intuitive way without the need to resort to involved mathematics. This will be done, as in this chapter, in a step-by-step fashion whereby new concepts will be presented as we progress in this familiarization process with the wireless propagation channel.

**References**

[Tranter] William H. Tranter, K. Sam Shanmugan, Theodore S. Rappaport, and Kurt L. Kosbar. Principles of Communication Systems Simulation with Wireless Applications, Prentice Hall, Professional Technical Reference, 2004, IBSN: 0-13-494790-8.

### [ITU] RECOMMENDATION ITU-R P.1057-1. Probability distributions relevant to radiowave propagation modelling. (1994-2001). International Telecommunication Union. Radiocommunication Sector

[HERFON] J.M. Hernando & F. Pérez-Fontán. *An Introduction to Mobile Communications Engineering*. Artech House, 1999.

[Lee] W.C.Y. Lee. *Mobile Communications Design Fundamentals. Wiley Series in Telecommunications and Signal Processing*. John Wiley & Sons, Ltd, Chichester, UK, 1993.

[Hata] M. Hata. Empirical formula for propagation loss in land mobile radio services. *IEEE Trans. Veh. Tech*., **29**(3), 1980, 317–325.

[Okumura]

[LeonG] A. Leon-Garcia. *Probability and Random Processes for Electrical Engineering*, Second Edition (International Edition). Addison-Wesley, 1994.

[Sklar] B. Sklar. *Rayleigh Fading Channels. Mobile Communications Handbook* (Ed. S.S. Suthersan). CRC Press, 1999.

[Suzuki],

[wikipedia] http://en.wikipedia.org/wiki/Chi-squared\_distribution (20131002)

[Gree99] L. J. Greenstein, D. G. Michelson, and V. Erceg. Moment-Method Estimation of the Ricean -Factor. IEEE COMMUNICATIONS LETTERS, VOL. 3, NO. 6, JUNE 1999 175-176

[APPLIED] S.C. Chapra. Applied Numerical Methods with MATLAB for Engineers and Scientists. 3rd edition. McGraw-Hill, 2012

[regression] R.C. Rumpf. Computational Science: Computational Methods in Engineering. Curve Fitting to Polynomials & Interpolation/Extrapolation. https://empossible.net/academics/emp4301\_5301/ (accessed 20200114)

Spring 2018. University of Texas at El Paso. College of Engineering. Department of Electrical and Computer Engineering.

[1st edition]

**Software Supplied**

In this section, we provide a list of functions and scripts, developed in MATLAB®, implementing the various projects and theoretical introductions mentioned in this chapter. They are the following:

|  |  |
| --- | --- |
| Specific functions |  |
| **RayleighCDFrms\_1** |  |
| **RayleighCDF** |  |
| **Rayleighpdf** |  |
| **GaussianCDF** |  |
|  | **~~BINSequalprobRayleigh~~** |
| **RayleighHIST** |  |
| **genRiceUncorrSeries** |  |
| **genRiceUncorrSeriesSNR** |  |
|  |  |
|  |  |
|  |  |
| General functions |  |
| **fCDF** |  |
| **fCDFbins** |  |
| Scripts |  |
| **Rayleigh\_pdf\_cdf** |  |
| fitRayleigh |  |
| fitRayChiTest |  |
| **fitSuzuki** |  |
| **fitRiceGreenstein** |  |
| **fitRiceGreenstein2** |  |
| **fitRiceGreensteinSNR** |  |
|  |  |
|  |  |
| **pathLossRegr** |  |
| **genChiCCDFs** |  |
| **genChiPDFs** |  |

Additionally, the following time series are supplied:

|  |
| --- |
| **RayleighSeries** |
| **SuzukiSeries** |
| **RiceSeries** |
| **RiceSeries2** |
| **RiceSeries3** |
| **pathloss** |

**ANNEX I.4.1 The Chi-square distribution and goodness of fit test**

This test involves the validation of a hypothesis, , the null hypothesis, which states that the distribution we are testing fits the data. The alternative hypothesis, , states that the chosen distribution does for fit the data.

The test is based on the central limit theorem where instead of having numerous independent random variables which, when added together, yield a Gaussian distribution, what we have is multiple random variables squared. In this case the resulting random variable converges to a Chi-Square distribution.

Assuming independent, standardized (zero mean and unit standard deviation) normal random variables, their sum of their squares is given by

|  |  |
| --- | --- |
|  | (xxx) |

The resulting random variable follows a **chi-squared distribution** with **degrees of freedom**, DoF. Its pdf is given by

|  |  |
| --- | --- |
|  | (xxx) |

where is a positive integer and is the Gamma function. The chi-square distribution is associated or is a special case of the Gamma distribution given below [wikipedia],

|  |  |
| --- | --- |
|  | (A.13) |

when , is a positive integer, and ,

is the Gamma function, that is,

|  |  |
| --- | --- |
|  | (xxx) |

The shape of the chi-square pdf as a function of its parameter , is presented in Figure 1.

|  |  |
| --- | --- |
|  |  |
| Figure 1. Representation of various chi-square pdfs as a function of the DoF (using script **genChiPDFs**) | Figure 2. Representation of various chi-square CDFs as a function of the DoF (using script **genChiCCDFs**) |

The **chi-squared goodness of fit test** verifies whether a data set fits a given theoretical distribution. The data is split into **mutually exclusive** events.

Four steps are followed in the chi-square test:

1.- We start off with a sample, , that is, the various measurements (in our example, the experimental time series). We partition the sample space, , into the union of disjoint intervals/bins.

2.- Then we compute the probability, , that an outcome falls in the -th interval under the assumption that the data follows the proposed distribution. This information is obtained from the assumed theoretical pdf, , as

|  |  |
| --- | --- |
|  | (xxx) |

where and are the lower and upper limit of interval/bin [Wikipedia]. Thus, if we have repetitions of the experiment (number of samples in our time-series), the expected number of outcomes in the -th interval would be ;

3.- Test parameter is defined as the averaged, weighted difference between the observed number of outcomes, , that fall in the -th interval (number of counts in the histogram in bin ), and the expected number, , i.e.,

|  |  |
| --- | --- |
|  | (xxx) |

where is the test statistic which should follow a chi-square distribution, is an observed frequency, is an expected (theoretical) frequency.

4.-The final step is verifying whether the test parameter, , is large or small in comparison with a threshold obtained from the chi-square distribution. It is clear that if the fit is good, then the value of should be small. The null hypothesis, i.e., that the measured data follows a given theoretical distribution, will be rejected if is too large, that is, if , where is the *threshold* for **significance level**, .

As said, the chi-square test is based on the fact that for large , the random variable follows a chi-square distribution with degrees of freedom. The threshold, , can be computed by finding the point at which , where is a chi-square random variable with *degrees of freedom*, DoF.

The thresholds for the 1% and 5% levels of significance and different degrees of freedom are given in Table I.4.2. The number of DoFs is , that is, the number of intervals or bins minus one. It is recommended that, if is the number of parameters extracted from the data (e.g., mean, standard deviation, etc.), then is better approximated by a chi-square distribution with degrees of freedom. Each estimated parameter decreases the degrees of freedom by one.

It is recommended that the expected number of outcomes in each interval be at least five or more. This will improve the accuracy of approximating the CCDF of by a chi-square distribution

**ANNEX I.4.x LINEAR LEAST-SQUARES REGRESSION** [APPLIED]

We want to develop a model that explains a measured data set. The first and obvious way to approach this problem is plotting the data in a so called scatter plot and observe the behavior of the data and may propose a curve that fits the data, see **Figure 14.8**. However, we need mathematical tools that performs this curve-fitting exercise in a quantitative way.

We first carry out the study by trying to fit a straight line to the data which is organized in observation pairs:

|  |  |
| --- | --- |
| , , ..., . | (xxx) |

The equation of a **straight line** is

|  |  |
| --- | --- |
|  | (14.8) |

where and are coefficients representing the intercept and the slope, respectively, and is the error, or ***residual****,* between the model and the observations.

We can solve for the error. We get

|  |  |
| --- | --- |
|  | (14.9) |

where we have the **residual** as the difference between

- the **true value** of and

- the **approximate value**, , predicted by the linear equation.

The strategy to be used is to **minimize the sum of the squares of the residuals**,

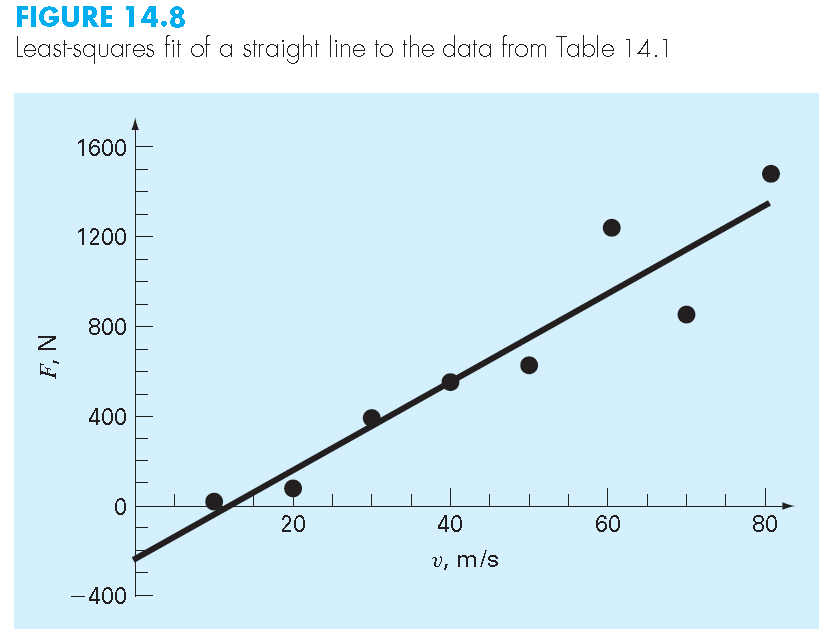
|  |  |
| --- | --- |
|  | (14.12) |

This criterion is called ***least squares***and has a number of advantages.

The straight line coefficients are given by the equations below,

|  |  |
| --- | --- |
| and | (14.15 and .16) |

with and being the means of and . Any line other than the one computed with the above equations results in a larger sum of the squares of the residuals. Thus, the line is unique and, in terms of our chosen criterion, is a "best" line through the points.



**Figure 14.8** Least squares fit of a straight line to the data from Table14.1

We can now try to quantify how well the straight line model just developed explains the observations. We have first **the sum of the squares of residuals**, , which measures the deviation of the data from the fitted line. We also can define another parameter

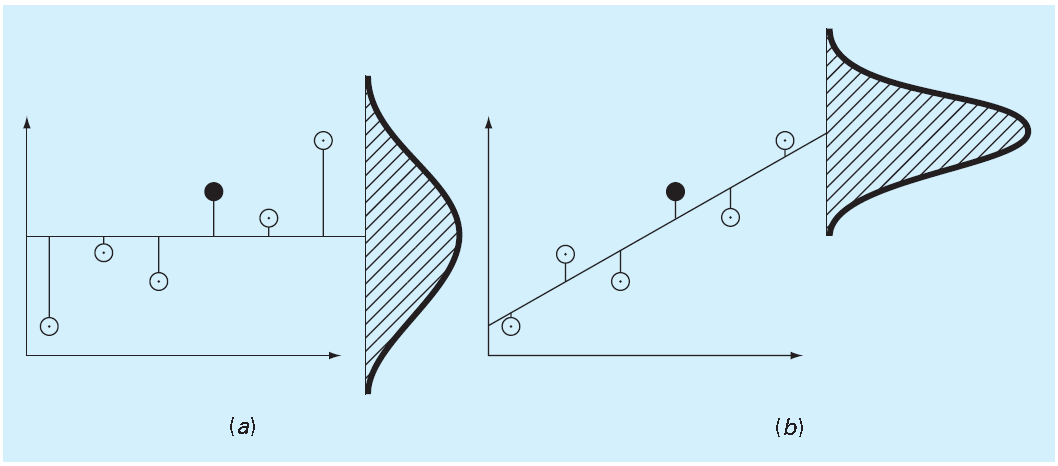
|  |  |
| --- | --- |
|  | (14.18) |

which measures the deviation of the data with respect to the data mean, a sort of variance. The actual standard deviations, from the fitted line and from the mean, require dividing by where the sample size, and further taking the square root.

For the case of the regression line obtained this parameter is called the ***standard error of the estimate****. and is defined as* [APPLIED]

|  |  |
| --- | --- |
|  | (14.19) |

Just as was the case with the standard deviation, the standard error of the estimate quantifies the spread of the data. However, quantifies the **spread *around the regression line***as shown in Fig. 14.10*b* in contrast to the standard deviation that quantified the **spread *around the mean***(Fig. 14.10*a*).



**FIGURE 14.10.** Regression data showing (*a*) the spread of the data around the mean of the dependent variable and (*b*) the spread of the data around the best-fit line. The reduction in the spread in going from (*a*) to (*b*), as indicated by the bell-shaped curves at the right, represents the improvement due to linear regression.

After performing the regression, we can compute , the sum of the squares of the residuals around the regression line with Eq. (14.12). This **characterizes the residual error that remains after the regression**.

It is, therefore, sometimes called the unexplained sum of the squares.

The difference between the two quantities, , quantifies **the improvement or error reduction** due to describing the data in terms of a straight line rather than as an average value. Because the magnitude of this quantity is scale-dependent, the difference is normalized to to yield

|  |  |
| --- | --- |
|  | (14.20) |

where is called the ***coefficient of determination***and is the ***correlation coefficient***().

For a perfect fit, and , signifying that the line explains 100% of the variability of the data.

For *,* and the fit represents no improvement.

**Multiple linear regression and polynomial regression**

in case they become necessary we provide here the least squares expressions in matrix form for computing the regression coefficients for two cases. One is the case where, instead of a simple line, a first degree polynomial, we want to develop a more complex model, for example a parabola or a third degree polynomial. In this case we are trying to fit a polynomial

|  |  |
| --- | --- |
|  | (xxx) |

If we have more measured data points, , larger than the order of the polynomial we are trying to fit, , then we can write the following equation system,

|  |  |
| --- | --- |
|  |  |

which can be put in matrix form, i.e.,

|  |  |
| --- | --- |
|  |  |

This equation can be put in shorthand notation as

|  |  |
| --- | --- |
|  |  |

The least squares estimation of the polynomial coefficients is given by

|  |  |
| --- | --- |
| **---> está bien ::::::**  **---------NO ME SALE ????? LO MISMO QUE** x = A\B |  |

This equation collapses to whenthe number of data points is the same as the sought for polynomial order, .

x = A\B

x = mldivide(A,B)

A similar procedure can be used in the case of multiple linear regression, that is, we have a solution depending on several measured pieces of information, i.e.,

|  |  |
| --- | --- |
|  |  |

Again, we can set a system of equations where we have data points, e.g. for sample we have

|  |  |
| --- | --- |
|  |  |

Setting the set of equations in matrix form, we have

|  |  |
| --- | --- |
|  |  |

which can be solved for the model coefficients in the same way as the previous case. If the model we want has an additional term, , that is our model is

Then, the set of equations can be put such that

~~In the bivariate case, we can solve this equation easily as follows,~~