

International Telecommunication Union

ITU-R
Radiocommunication Sector of ITU

Recommendation ITU-R P.1057-3
(09/2013)

**Probability distributions relevant to
radiowave propagation modelling**

P Series
Radiowave propagation

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from <http://www.itu.int/ITU-R/go/patents/en> where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Recommendations

(Also available online at <http://www.itu.int/publ/R-REC/en>)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	Broadcasting service (television)
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

Note: This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.

Electronic Publication
Geneva, 2013

© ITU 2013

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without written permission of ITU.

RECOMMENDATION ITU-R P.1057-3

Probability distributions relevant to radiowave propagation modelling

(1994-2001-2007-2013)

Scope

This Recommendation describes the various probability distributions relevant to radiowave propagation modelling and predictions.

The ITU Radiocommunication Assembly,

considering

- a) that the propagation of radio waves is mainly associated with a random medium which makes it necessary to analyse propagation phenomena by means of statistical methods;
- b) that, in most cases, it is possible to describe satisfactorily the variations in time and space of propagation parameters by known statistical distributions;
- c) that it is therefore important to know the fundamental properties of the probability distributions most commonly used in statistical propagation studies,

recommends

- 1 that the statistical information relevant to propagation modelling provided in Annex 1 should be used in the planning of radiocommunication services and the prediction of system performance parameters;
- 2 that the step-by-step procedure provided in Annex 2 should be used to approximate a complementary cumulative distribution by a log-normal complementary cumulative distribution.

Annex 1**Probability distributions relevant to radiowave propagation modelling****1 Introduction**

Experience has shown that information on the mean values of the signals received is not sufficient to characterize the performance of radiocommunication systems. The variations in time, space and frequency also have to be taken into consideration.

The dynamic behaviour of both wanted signals and interference plays a decisive role in the analysis of system reliability and in the choice of system parameters such as modulation type. It is essential to know the extent and rapidity of signal fluctuations in order to be able to specify such parameters as type of modulation, transmit power, protection ratio against interference, diversity measures, coding method, etc.

For the description of communication system performance it is often sufficient to observe the time series of signal fluctuation and characterize these fluctuations as a stochastic process. Modelling of signal fluctuations for the purpose of predicting radio system performance, however, requires also knowledge of the mechanisms of interaction of radio waves with the atmosphere (neutral atmosphere and the ionosphere).

The composition and physical state of the atmosphere is highly variable in space and time. Wave interaction modelling, therefore, requires extensive use of statistical methods to characterize various physical parameters describing the atmosphere as well as electrical parameters defining signal behaviour and the interaction processes via which these parameters are related.

In the following, some general information is given on the most important probability distributions. This may provide a common background to the statistical methods for propagation prediction used in the Recommendations of the Radiocommunication Study Groups.

2 Probability distributions

Stochastic processes are generally described either by a probability density function or by a cumulative distribution function. The probability density function, here denoted by $p(x)$ for the variable x , is such that the probability of x taking a value in the infinitesimal interval x to $x + dx$ is $p(x) dx$. The cumulative distribution function, denoted by $F(x)$, gives the probability that the variable takes a value less than x , i.e. the functions are related as follows:

$$p(x) = \frac{d}{dx} [F(x)]$$

or:

$$F(x) = \int_c^x p(t) dt$$

where

c is the lowest limit of the values which t can take.

The following distributions are the most important:

- normal or Gaussian distribution;
- log-normal distribution;
- Rayleigh distribution;
- combined log-normal and Rayleigh distribution;
- Nakagami-Rice distribution (Nakagami n -distribution);
- gamma distribution and exponential distribution;
- Nakagami m -distribution;
- Pearson χ^2 distribution.

3 Normal distribution

This distribution is applied to a continuous variable of any sign. The probability density is of the type:

$$p(x) = e^{-T(x)} \quad (1)$$

$T(x)$ being a non-negative second degree polynomial. If as parameters we use the mean, m , and the standard deviation, σ , then $p(x)$ is written in the usual way:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] \quad (2)$$

hence:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2\right] dt = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-m}{\sigma\sqrt{2}}\right)\right] \quad (3)$$

with:

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad (4)$$

The solid lines in Fig. 1 represent the functions $p(x)$ and $F(x)$ with m equal to zero and σ equal to unity. The cumulative normal distribution $F(x)$ is generally tabulated in a short form for the same conditions. Table 1 gives the correspondence between x and $F(x)$ for a number of round values of x or $F(x)$.

TABLE 1

x	$1 - F(x)$	x	$1 - F(x)$
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}

For the purpose of practical calculations, $F(x)$ can be represented by approximate functions, for example the following which is valid for positive x with a relative error less than 2.8×10^{-3} :

$$1 - F(x) = \frac{\exp(-x^2/2)}{\sqrt{2\pi} \left(0.661x + 0.339\sqrt{x^2 + 5.51}\right)} \quad (5)$$

A normal distribution is mainly encountered when values of the quantity considered result from the additive effect of numerous random causes, each of them of relatively slight importance.

In propagation most of the physical quantities involved (power, voltage, fading time, etc.) are essentially positive quantities and cannot therefore be represented directly by a normal distribution. On the other hand this distribution is used in two important cases:

- to represent the fluctuations of a quantity around its mean value (scintillation);
- to represent the logarithm of a quantity. We then obtain the log-normal distribution which is studied later.

Diagrams in which one of the coordinates is a so-called normal coordinate are available commercially, i.e. the graduation is such that a normal distribution is represented by a straight line. These diagrams are very frequently used even for the representation of non-normal distributions.

4 Log-normal distribution

This is the distribution of a positive variable whose logarithm has a normal distribution. It is possible therefore to write directly the probability density and the cumulative density:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp \left[-\frac{1}{2} \left(\frac{\ln x - m}{\sigma} \right)^2 \right] \quad (6)$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{1}{t} \exp \left[-\frac{1}{2} \left(\frac{\ln t - m}{\sigma} \right)^2 \right] dt = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\ln x - m}{\sigma\sqrt{2}} \right) \right] \quad (7)$$

However, in these relations m and σ are the mean and the standard deviation not of the variable x but of the logarithm of this variable.

The log-normal distribution is very often found in connection with propagation, mainly for quantities associated either with a power or field-strength level or a time. Power or field-strength levels are generally only expressed in decibels so that sometimes reference is made to a log-normal distribution simply as a normal distribution. This usage is not recommended. In the case of time (for example fading durations), the log-normal distribution is always used explicitly because the natural variable is the second or the minute and not their logarithm.

Since the reciprocal of a variable with a log-normal distribution also has a log-normal distribution, this distribution is sometimes found in the case of rates (reciprocals of time). For example, it is used to represent rainfall rate distributions.

In comparison with a normal distribution, it can be considered that a log-normal distribution means that the numerical values of the variable are the result of the action of numerous causes of slight individual importance which are multiplicative.

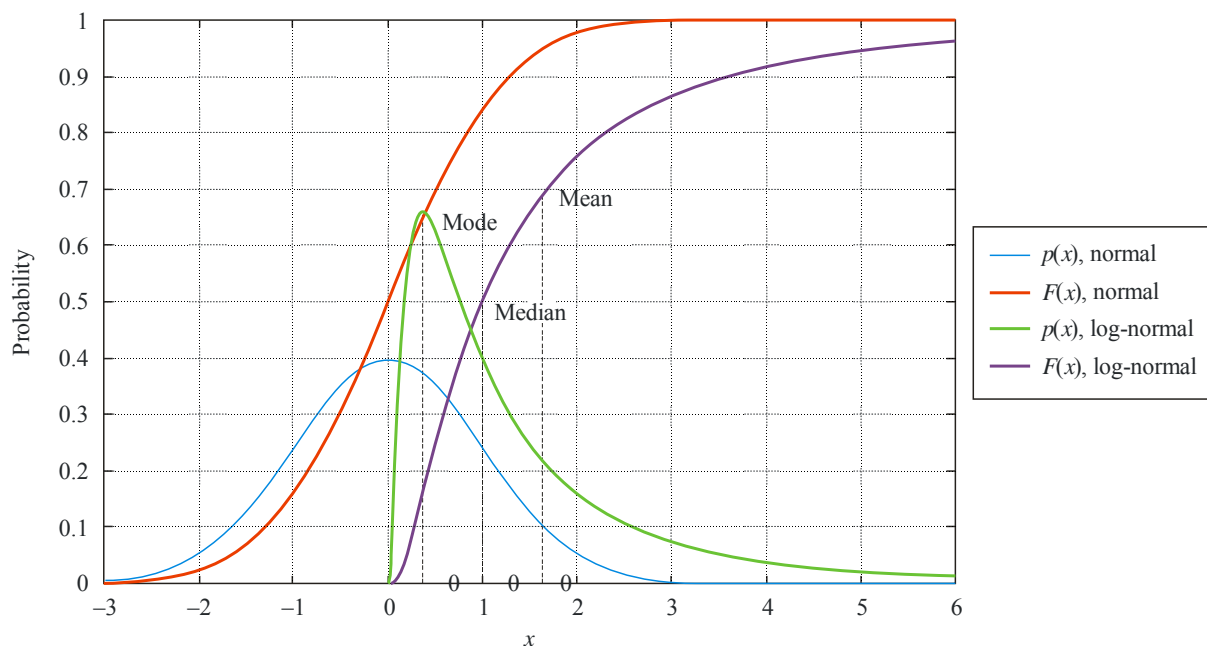
When considered in numerical terms, a log-normal distribution is extremely asymmetrical, unlike the normal distribution. In particular, the mean value, the median value and the most probable value (often called the mode) are not identical (see the dashed lines in Fig. 1).

The characteristic quantities of the numerical variable x are:

- most probable value: $\exp(m - \sigma^2)$;
- median value: $\exp(m)$;
- mean value: $\exp\left(m + \frac{\sigma^2}{2}\right)$;

- root mean square value: $\exp(m + \sigma^2)$;
- standard deviation: $\exp\left(m + \frac{\sigma^2}{2}\right) \sqrt{\exp(\sigma^2) - 1}$.

FIGURE 1
Normal and log normal distributions



P.1057-01

5 Rayleigh distribution

The Rayleigh distribution applies to a positive continuous variable. It is linked with the normal distribution as follows. Given a two-dimensional normal distribution with two independent variables y and z of mean zero and the same standard deviation σ , the random variable

$$x = \sqrt{y^2 + z^2} \tag{8}$$

has a Rayleigh distribution. The most probable value of x is equal to σ . The Rayleigh distribution represents the distribution of the length of a vector which is the sum of a large number of vectors of similar amplitudes whose phases have a uniform distribution.

The probability density and the cumulative distribution are given by:

$$p(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{9}$$

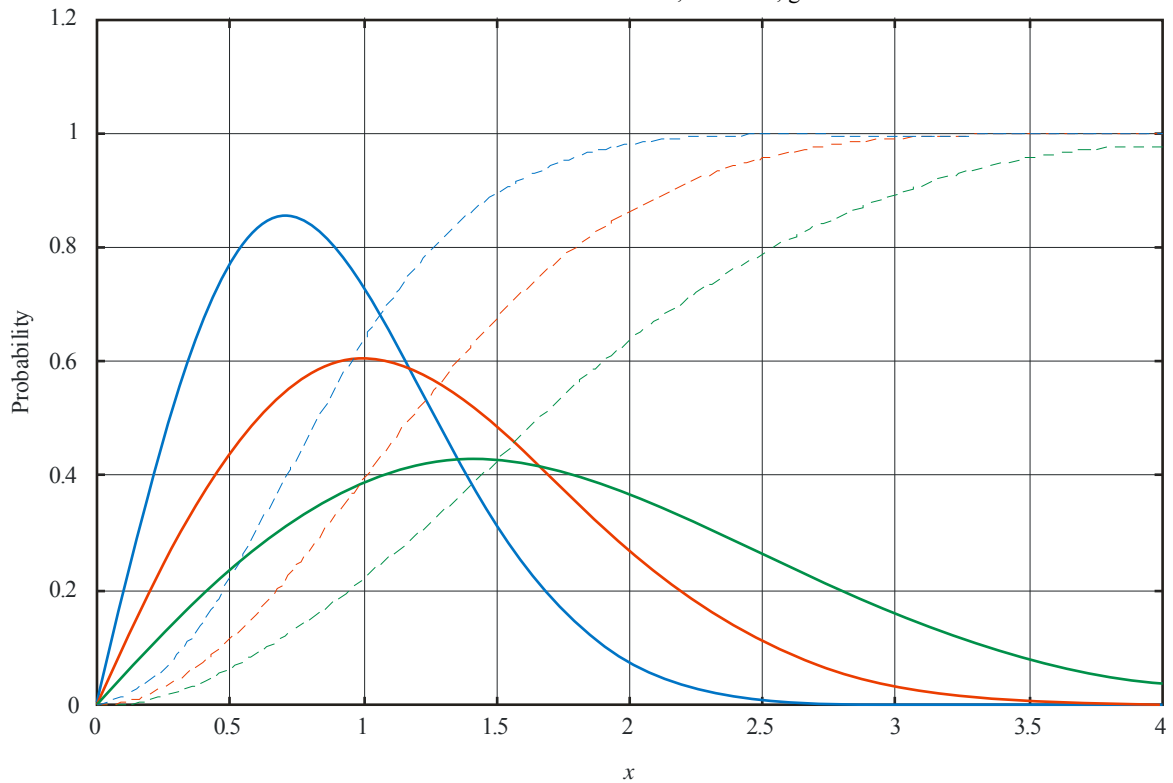
$$F(x) = 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{10}$$

Figure 2 gives examples of these functions $p(x)$ and $F(x)$ for three different values of b .

FIGURE 2

Rayleigh distribution

$p(x)$ is shown as solid lines and $F(x)$ as dashed lines
for three different values of b : blue $b = 1$; red $b = 2$; green $b = 4$



P.1057-02

The characteristic values of the variable are as follows:

- most probable value: $\frac{b}{\sqrt{2}}$;
- median value: $\frac{b}{2} \ln 2 \approx 0.833b$;
- mean value: $\frac{b}{2} \sqrt{\pi} \approx 0.886b$;
- root mean square value: b ;
- standard deviation: $b \sqrt{1 - \pi/4} \approx 0.463b$.

The Rayleigh distribution is often only used near the origin, i.e. for low values of x . In this case we have:

$$F(x) \approx \frac{x^2}{b^2} \quad (11)$$

This expression can be interpreted as follows: the probability that the random variable X will have a value of less than x is proportional to the square of this value. If the variable in question is a voltage, its square represents the power of the signal. In other words, on a decibel scale the power decreases by 10 dB for each decade of probability. This property is often used to find out whether a received level has a Rayleigh distribution at least asymptotically. It should be noted, however, that other distributions can have the same behaviour.

In particular the Rayleigh distribution occurs for scatter from independent, randomly-located scatterers for which no scattering component dominates.

Footnote: $b = \sigma \cdot \sqrt{2}$

6 Combined log-normal and Rayleigh distribution

In some cases the distribution of a random variable can be regarded as the resultant of a combination of two distributions, i.e. a log-normal distribution for long-term variations and a Rayleigh distribution for short-term variations. The distribution of instantaneous values is obtained by considering a Rayleigh variable whose mean (or mean square) value is itself a random variable having a log-normal distribution.

If m and σ are used to designate the mean and the standard deviation of the normal distribution associated with the log-normal distribution, the following distribution is obtained:

$$1 - F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left[-x^2 e^{-2\sigma u} - \frac{u^2}{2}\right] du \quad (12)$$

In this formula the standard deviation σ is expressed in nepers. If σ' is used to designate its value in decibels, we have:

$$\sigma = 0.115 \sigma' \quad (13)$$

Figure 3 shows a graph of this distribution for a number of values of the standard deviation, the value of m being taken to be equal to zero.

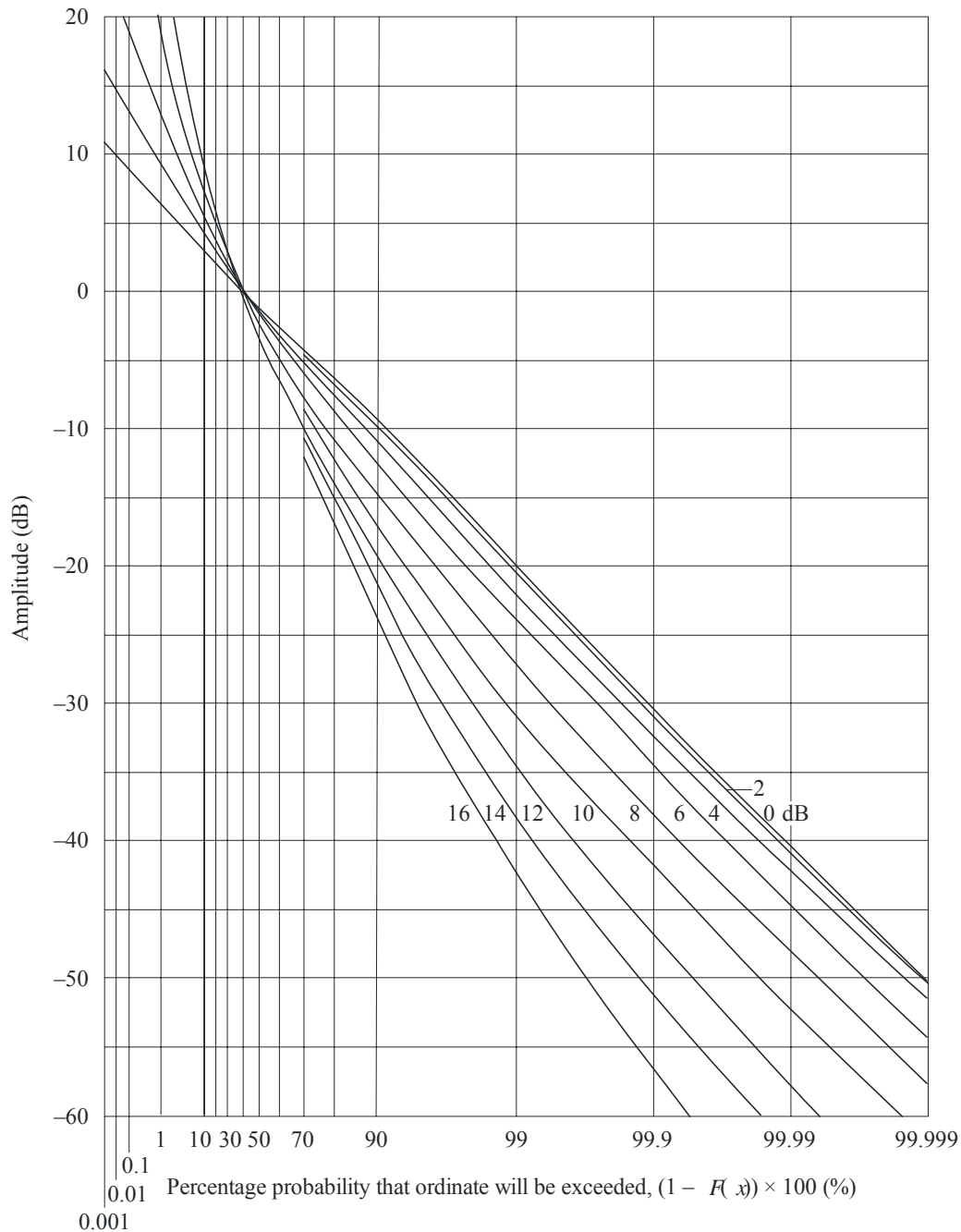
The characteristic values of the variable x are as follows:

- most probable value of x : $\ln\left(\frac{1}{\sqrt{2}}\right)$;
- median value of x : $\ln(\sqrt{\ln 2})$;
- mean value of x : $\ln\left(\frac{\sqrt{\pi}}{2}\right)$;
- root mean square value of x : 0;
- standard deviation of x : $\ln\left(\frac{\sqrt{4-\pi}}{2}\right)$.

The distribution occurs mainly in propagation via inhomogeneities of the medium when the characteristics of the latter have non-negligible long-term variations, as for example in the case of tropospheric scatter.

FIGURE 3

Combined log-normal and Rayleigh distributions (with standard deviation of the log-normal distribution as parameter)



P.1057-03

7 Nakagami-Rice distribution (Nakagami n -distribution) (See Note 1)

NOTE 1 – Not to be confused with the Nakagami m -distribution.

The Nakagami-Rice distribution is also derived from the normal distribution and it generalizes the Rayleigh distribution. It may be considered as the distribution of the length of a vector which is the sum of a fixed vector and of a vector whose length has a Rayleigh distribution. Alternatively, given a two-dimensional normal distribution with two independent variables x and y and with the same standard deviation σ , the length of a vector joining a point in the distribution to a fixed point different from the centre of the distribution will have a Nakagami-Rice distribution.

If a is used to designate the length of the fixed vector and σ the most probable length of the Rayleigh vector, the probability density is given by:

$$p(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2 + a^2}{2\sigma^2}\right) I_0\left(\frac{ax}{\sigma^2}\right) \quad (14)$$

where I_0 is a modified Bessel function of the first kind and of zero order.

This distribution depends on two parameters but for the purposes of propagation problems it is necessary to choose a relation between the amplitude a of the fixed vector and the root mean square amplitude $\sigma\sqrt{2}$ of the random vector. This relation depends on the application envisaged. The two main applications are as follows:

- a) Power in the fixed vector is constant, but the total power in fixed and random components varies

For studies of the influence of a ray reflected by a rough surface, or for a consideration of multipath components in addition to a fixed component, the mean power is given by $(a^2 + 2\sigma^2)$. The distribution is often defined in terms of a parameter K :

$$K = 10 \log\left(\frac{a^2}{2\sigma^2}\right) \quad \text{dB} \quad (15)$$

which is the ratio of the powers in the fixed vector and the random component.

- b) Total power in the fixed and random components is constant, but both components vary

For the purpose of studying multipath propagation through the atmosphere, it can be considered that the sum of the power carried by the fixed vector and the mean power carried by the random vector is constant since the power carried by the random vector originates from that of the fixed vector. If the total power is taken to be unity, one then has:

$$a^2 + 2\sigma^2 = 1 \quad (16)$$

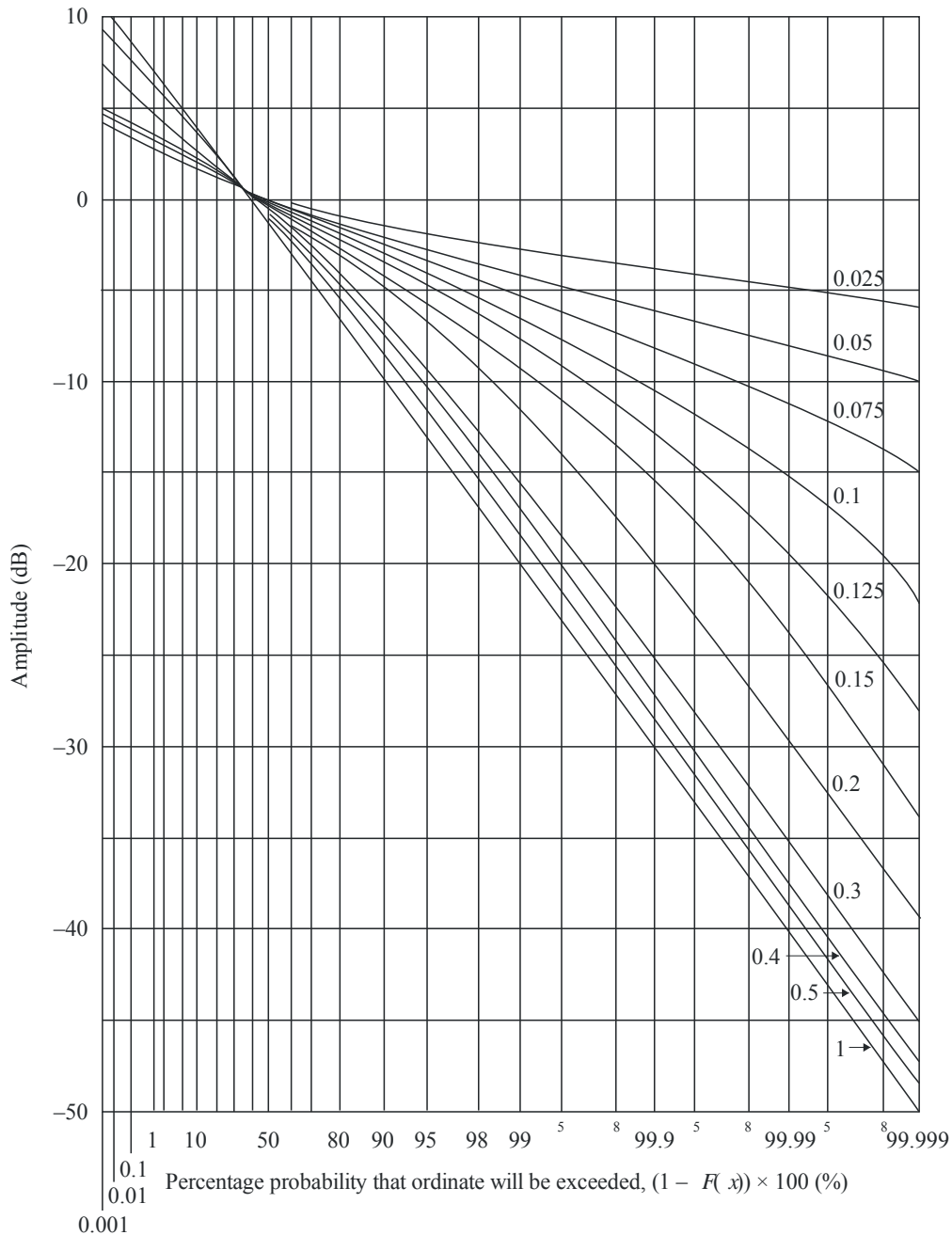
and the fraction of the total power carried by the random vector is then equal to $2\sigma^2$. If X is used to designate the instantaneous amplitude of the resultant vector and x a numerical value of this amplitude, we find that the probability of having an instantaneous level greater than x is given by:

$$\text{Prob}(X > x) = 1 - F(x) = 2 \exp\left(-\frac{a^2}{2\sigma^2}\right) \int_{x/\sigma\sqrt{2}}^{\infty} v \exp(-v^2) I_0\left(\frac{2va}{\sigma\sqrt{2}}\right) dv \quad (17)$$

Figure 4 shows this distribution for different values of the fraction of power carried by the random vector.

FIGURE 4

Nakagami-Rice distribution for a constant total power (with the fraction of power carried by the random vector as parameter)



P.1057-04

For the purpose of practical applications use has been made of a decibel scale for the amplitudes, and for the probabilities, a scale such that a Rayleigh distribution is represented by a straight line. It will be seen that for values of the fraction of power in the random vector above about 0.5, the curves approach a limit corresponding to a Rayleigh distribution. This is because in this case the fixed vector has an amplitude of the same order of magnitude as that of the random vector and it is practically indistinguishable from it. On the other hand for small values of this fraction it can be shown that the distribution of the amplitude tends towards a normal distribution.

While the amplitude has a Nakagami-Rice distribution, the probability density function of the phase is:

$$p(\theta) = \frac{1}{2\pi} \left\{ 1 + \sqrt{\frac{\pi}{2}} \frac{a \cos \theta}{\sigma} e^{-\frac{a^2 \cos^2 \theta}{2\sigma^2}} \left[1 + \operatorname{erf} \left(\frac{a \cos \theta}{\sqrt{2}\sigma} \right) \right] \right\} \cdot e^{-\frac{a^2}{2\sigma^2}} \quad (18)$$

where:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-t^2} dt \quad (19)$$

8 Gamma distribution and exponential distribution

Unlike the previous distributions which derive from a Gaussian distribution, the gamma distribution is derived from the exponential distribution of which it is a generalization. It is applied to a positive and non-limited variable. The probability density is:

$$p(x) = \frac{\alpha^v}{\Gamma(v)} x^{v-1} e^{-\alpha x} \quad (20)$$

where Γ is the Euler function of second order.

This distribution depends on two parameters α and v . However α is only a scale parameter of variable x . Characteristic values of the variable are:

- mean value: $\frac{v}{\alpha}$
- root mean square value: $\frac{\sqrt{v(1+v)}}{\alpha}$
- standard deviation: $\frac{\sqrt{v}}{\alpha}$

The integral expressing the cumulative distribution cannot be evaluated in closed form except for integral values of v . On the other hand the following expansions are possible:

Series approximation for $x \ll 1$:

$$F(x) = \frac{1}{\Gamma(v+1)} e^{-\alpha x} (\alpha x)^v \left[1 + \frac{\alpha x}{v+1} + \frac{(\alpha x)^2}{(v+1)(v+2)} + \dots \right] \quad (21)$$

Asymptotic approximation for $x \gg 1$:

$$1 - F(x) = \frac{1}{\Gamma(v)} e^{-\alpha x} (\alpha x)^{v-1} \left[1 + \frac{v-1}{\alpha x} + \frac{(v-1)(v-2)}{(\alpha x)^2} + \dots \right] \quad (22)$$

For ν equal to unity we find an exponential distribution. For integer ν the asymptotic expansion has a finite number of terms and gives the gamma distribution in an explicit form.

In propagation the useful values of ν are very low values of the order of 1×10^{-2} to 1×10^{-4} . For ν in the vicinity of zero, we have:

$$\frac{1}{\Gamma(\nu)} \simeq \frac{\nu}{\Gamma(\nu+1)} \simeq \nu \quad (23)$$

It is possible therefore to write for ν small and αx not too small:

$$1 - F(x) \simeq \nu \int_{\alpha x}^{\infty} \frac{e^{-t}}{t} dt \quad (24)$$

For practical calculations it is possible to find an approximation to the above integral, for example the following:

$$1 - F(x) \simeq \nu \frac{e^{-\alpha x}}{0.68 + \alpha x + 0.28 \log \alpha x} \quad (25)$$

which is valid for $\nu < 0.1$ and $\alpha x > 0.03$.

The cumulative distribution of the complementary gamma function for small values of ν is shown in Fig. 5. It can be seen that the probability of the variable x being significantly greater than zero is always small. In particular this explains the use of the gamma distribution to represent rainfall rates since the total percentage of rainfall time is generally of the order of 2 to 10%.

9 Nakagami m -distribution (see Note 1)

NOTE 1 – In this section m denotes a parameter of the Nakagami m -distribution; it is not a mean value as in previous sections of this Annex.

This distribution is applied to a non-limited positive variable. The probability density is equal to:

$$p(x) = \frac{2m^m}{\Gamma(m) \Omega^m} x^{2m-1} e^{-\frac{m}{\Omega} x^2} \quad (26)$$

Ω is a scale parameter equal to the mean value of x^2 .

$$\overline{x^2} = \Omega \quad (27)$$

This distribution has various relationships with the previous distributions:

- if a variable has a Nakagami m -distribution, the square of this variable has a gamma distribution;
- for $m = 1$ we obtain a Rayleigh distribution;
- for $m = 1/2$ we obtain the one-sided normal distribution.

The Nakagami m -distribution and the Nakagami-Rice distribution can thus be regarded as two different generalizations of the Rayleigh distribution. It should be noted that for very low signal levels, the slope of the Nakagami m -distribution tends towards a value which depends on the parameter m , unlike the Nakagami-Rice distribution for which the limit slope is always the same (10 dB per decade of probability). The cumulative Nakagami m -distribution for various values of parameter m is shown in Fig. 6.

FIGURE 5
Gamma distribution ($\alpha = 1, \nu \leq 0.1$)

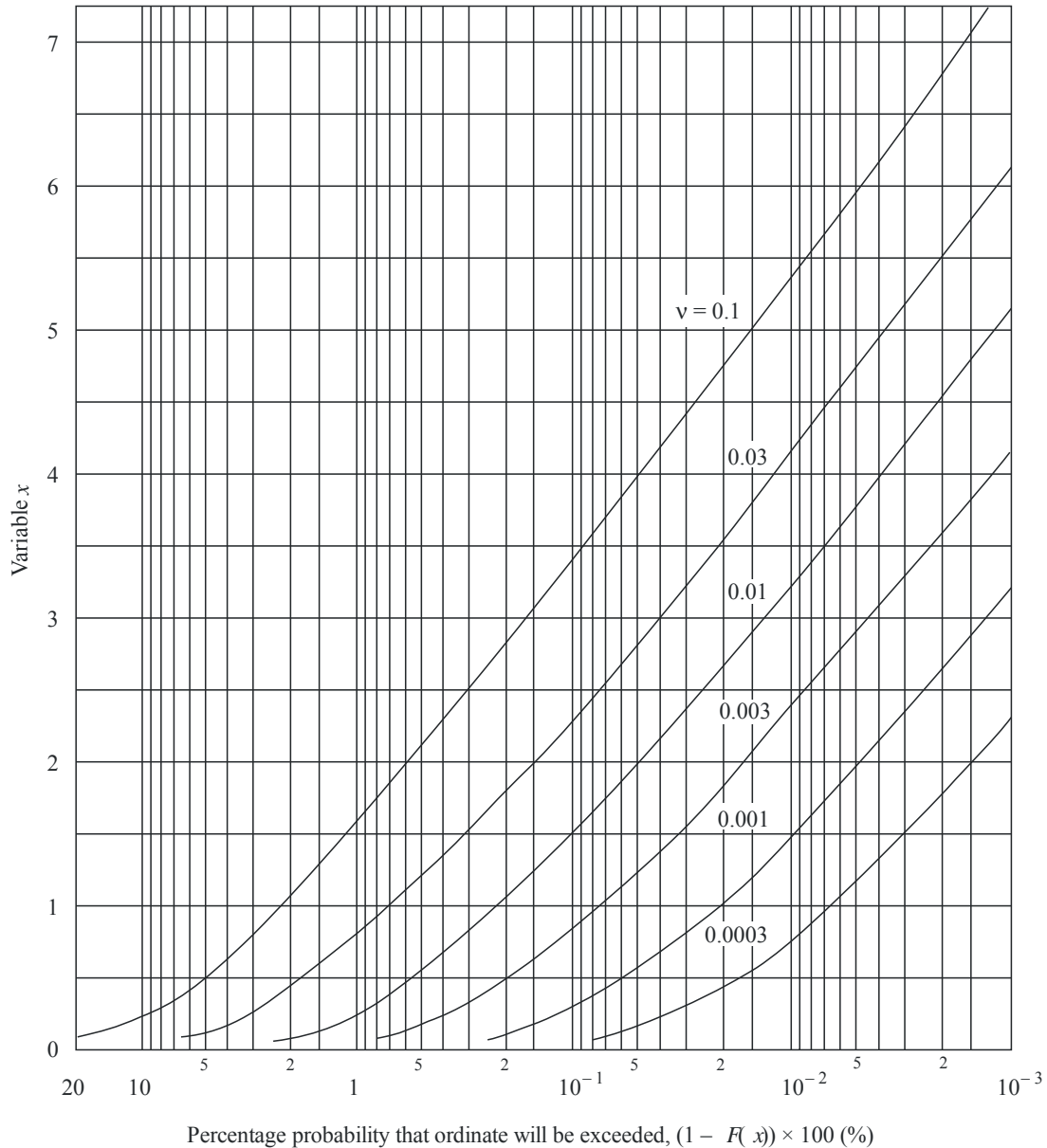
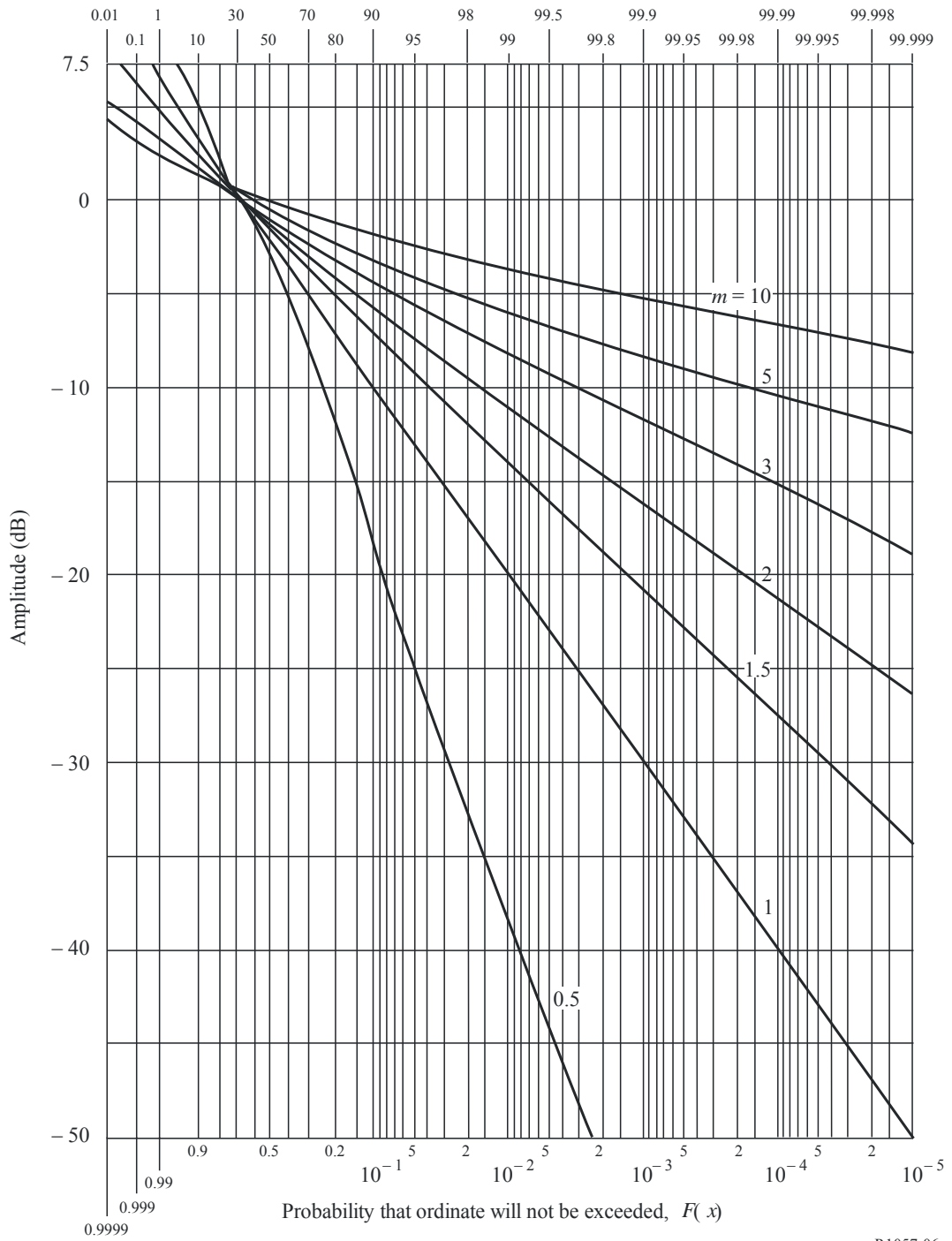


FIGURE 6

Nakagami- m distribution ($\overline{\chi^2} = 1$)

Percentage probability that ordinate will be exceeded, $(1 - F(x)) \times 100$ (%)



10 Pearson χ^2 distribution

The probability density is given by the equation:

$$p(\chi^2) = \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{v}{2}-1} \quad (28)$$

χ^2 is a non-limited positive variable and the parameter v , a positive integer, is termed the number of degrees of freedom of the distribution. Γ represents the Euler function of second order. Depending on the parity of v , one has

$$v \text{ even: } \Gamma\left(\frac{v}{2}\right) = \left(\frac{v}{2} - 1\right)! \quad (29)$$

$$v \text{ odd: } \Gamma\left(\frac{v}{2}\right) = \left(\frac{v}{2} - 1\right) \left(\frac{v}{2} - 2\right) \dots \frac{1}{2} \sqrt{\pi} \quad (30)$$

The cumulative distribution is given by:

$$F(\chi^2) = \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} \int_0^{\chi^2} e^{-\frac{t}{2}} t^{\frac{v}{2}-1} dt \quad (31)$$

The mean and standard deviation are given by:

$$m = v \quad (32)$$

$$\sigma = \sqrt{2v} \quad (33)$$

An essential property of the χ^2 distribution is that, if n variables x_i have Gaussian distributions of mean m_i and standard deviation σ_i , the variable:

$$\sum_{i=1}^n \left(\frac{x_i - m_i}{\sigma_i} \right)^2 \quad (34)$$

has a χ^2 distribution of n degrees of freedom. In particular, the square of a small Gaussian variable has a χ^2 distribution of one degree of freedom.

If several independent variables have χ^2 distributions, their sum also has a χ^2 distribution with a number of degrees of freedom equal to the sum of the degrees of freedom of each variable.

The χ^2 distribution is not fundamentally different from the gamma distribution. Conversion from one to the other can be effected by the equations:

$$\frac{\chi^2}{2} = \alpha x \quad (35)$$

$$\frac{\nu}{2} = n \quad (36)$$

Similarly, conversion from χ^2 distribution to the Nakagami- m distribution can be achieved by:

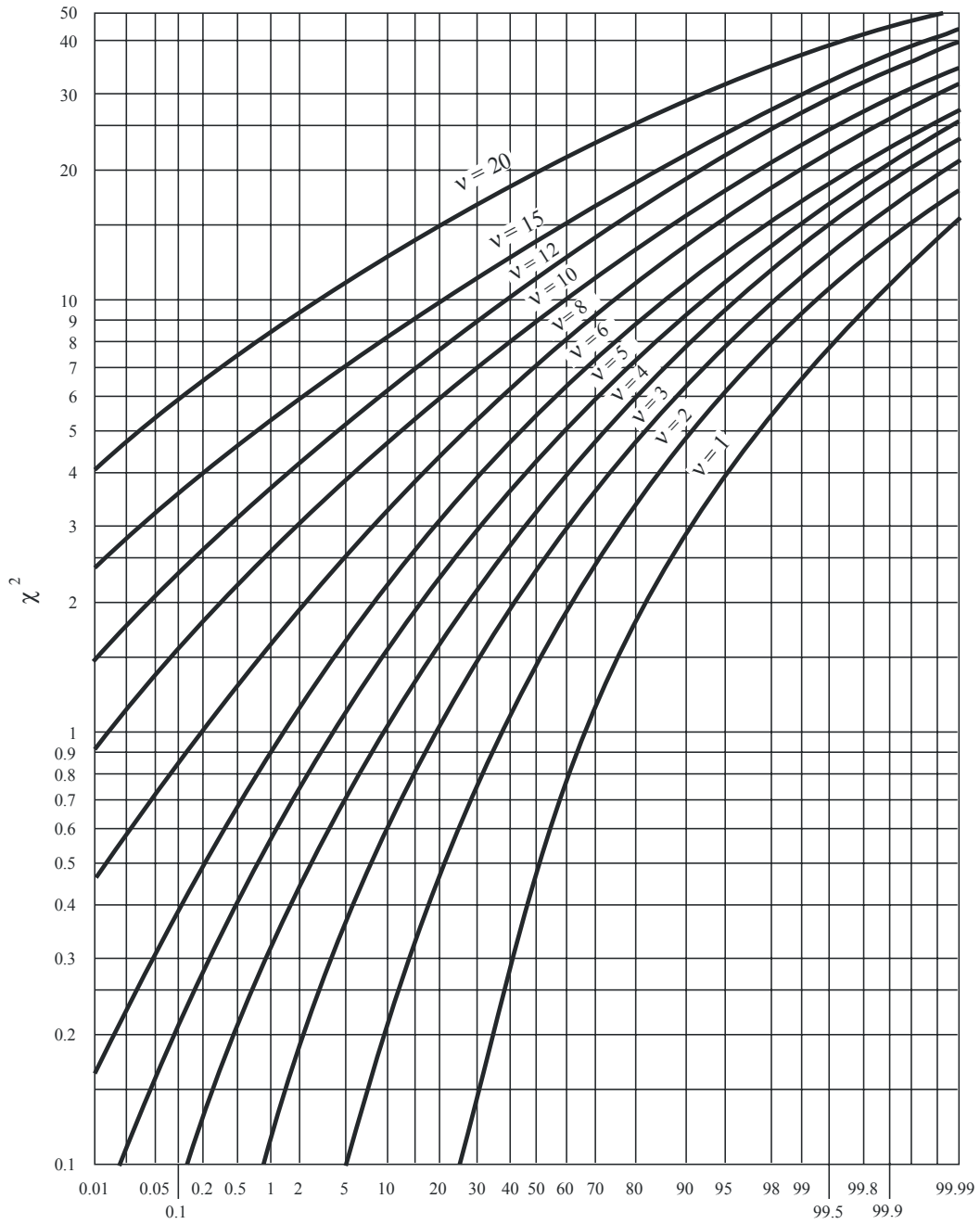
$$\frac{\chi^2}{2} = \frac{m}{\Omega} x^2 \quad (37)$$

$$\frac{\nu}{2} = m \quad (38)$$

The χ^2 distribution is used in statistical tests to determine whether a set of experimental values of a quantity (rainfall rate, attenuation, etc.) can be modelled by a given statistical distribution.

Figure 7 gives a graphic representation of the distribution for a number of values of ν .

FIGURE 7
 χ^2 distribution



Percentage probability that ordinate will not be exceeded, $H(\chi^2) \times 100\%$

Annex 2

Step-by-step procedure to approximate a complementary cumulative distribution by a log-normal complementary cumulative distribution

1 Background

The log-normal cumulative distribution is defined as:

$$\begin{aligned}
 F(x) &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{1}{t} \exp\left[-\frac{1}{2}\left(\frac{\ln t - m}{\sigma}\right)^2\right] dt \\
 &= \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{\ln x - m}{\sigma\sqrt{2}}\right) \right]
 \end{aligned} \tag{39}$$

or equivalently:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\ln x - m}{\sigma}} \exp\left[-\frac{t^2}{2}\right] dt \tag{40}$$

Similarly, the log-normal complementary cumulative distribution is defined as:

$$\begin{aligned}
 G(x) &= \frac{1}{\sigma\sqrt{2\pi}} \int_x^{\infty} \frac{1}{t} \exp\left[-\frac{1}{2}\left(\frac{\ln t - m}{\sigma}\right)^2\right] dt \\
 &= \frac{1}{2} \operatorname{erfc}\left(\frac{\ln x - m}{\sigma\sqrt{2}}\right) = \frac{1}{2} \left[1 - \operatorname{erf}\left(\frac{\ln x - m}{\sigma\sqrt{2}}\right) \right]
 \end{aligned} \tag{41}$$

or equivalently:

$$\begin{aligned}
 G(x) &= \frac{1}{\sqrt{2\pi}} \int_{\frac{\ln x - m}{\sigma}}^{\infty} \exp\left[-\frac{t^2}{2}\right] dt \\
 &= Q\left(\frac{\ln x - m}{\sigma}\right)
 \end{aligned} \tag{42}$$

where $Q(\cdot)$ is the normal complementary cumulative probability integral. The parameters m and σ can be estimated from a set of n pairs (G_i, x_i) as described in the following paragraph.

2 Procedure

Estimate the two log-normal parameters m and σ as follows:

Step 1: Construct the set of n pairs (G_i, x_i) where G_i is the probability that x_i is exceeded.

Step 2: Transform the set of n pairs from (G_i, x_i) to $(Z_i, \ln x_i)$ where:

$$Z_i = \sqrt{2} \operatorname{erfc}^{-1}(2G_i) = \sqrt{2} \operatorname{erf}^{-1}(1-2G_i) \text{ or, equivalently, } Z_i = Q^{-1}(G_i)$$

Step 3: Determine the variables m and σ by performing a least squares fit to the linear function:

$$\ln x_i = \sigma Z_i + m$$

as follows:

$$\sigma = \frac{n \sum_{i=1}^n Z_i \ln x_i - \sum_{i=1}^n Z_i \sum_{i=1}^n \ln x_i}{n \sum_{i=1}^n Z_i^2 - \left[\sum_{i=1}^n Z_i \right]^2}$$

$$m = \frac{\sum_{i=1}^n \ln x_i - \sigma \sum_{i=1}^n Z_i}{n}$$
