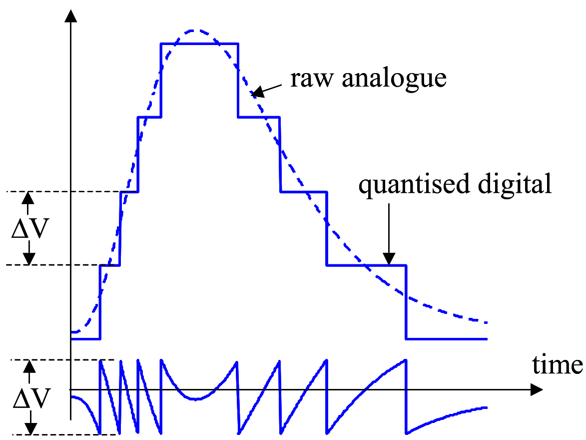
1 WYNTKA… Statistical Distributions

A few statistical distributions turn up over and over again in communications engineering, including the uniform, Geometric, negative exponential, Poisson, Gaussian and Rayleigh. This chapter is about these common distributions: where they come from, why they are useful, and what properties they have. I'll start with the simplest continuous distribution (that is, one in which the stochastic variable is not restricted to taking on certain discrete values): the uniform distribution.

1.1 The Uniform Distribution

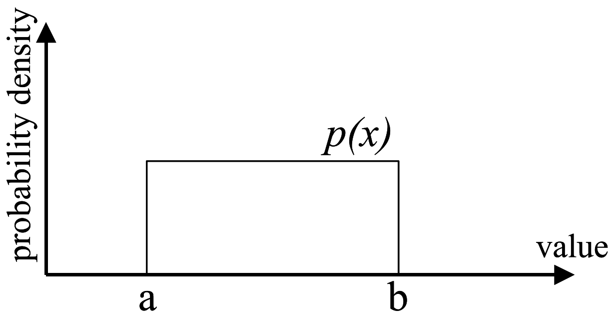
There's not much to say about the uniform distribution. Not many physical processes in real life are friendly enough to obey such a simple distribution, although there are a couple of notable cases where this distribution does find some use. One such case occurs when you have no idea which direction a radio wave might be arriving from – and in this case the simplest choice is to use a uniform distribution with a value of $1/2\pi$ between 0 and 2π to represent the unknown angle.

The other common case is in the noise generated by an analogue-to-digital converter which rounds up the input to the nearest of the discrete levels which it uses to represent values. The error introduced by this *quantisation* process has a uniform distribution from $-\Delta V/2$ to $\Delta V/2$ where ΔV is the distance between two adjacent digital (output) levels.



Another interesting thing about the uniform distribution is that most random number generators generate uniformly-distributed random numbers, so it's usually the place to start from to generate random numbers with any particular distribution.

With the uniform distribution, the probability of the random variable taking any particular value within a certain given range (in the figure below, a range from *a* to *b*) is a constant.



The properties of the uniform distribution are summarised in the following table:

The Uniform Distribution	
Probability Distribution: $p_x(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \le x < b \\ 0 & x > b \end{cases}$	
Cumulative Probability Distribution: $P_x(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & x > b \end{cases}$	
Mean: $\frac{a+b}{2}$	Mode: Any value between *a* and *b* is a mode of this distribution
Median: $\frac{a+b}{2}$	Standard Deviation: $\frac{(b-a)}{2\sqrt{3}}$
Variance: $\frac{(b-a)^2}{12}$	rms: $\sqrt{\frac{a^2+ab+b^2}{3}}$

1.2 The Gaussian (Normal) Distribution

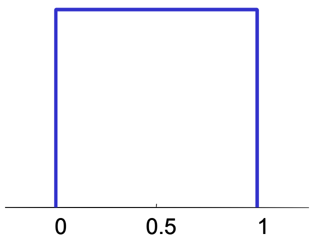
At the other extreme of usefulness, the Gaussian or normal distribution is without much doubt the most useful and most common distribution in engineering. The very wide-range of uses of the Gaussian distribution is due to the Central Limit Theorem, which states that the distribution of the sum of a large number of independent stochastic variables, each of which has the same probability distribution[1], tends towards a normal distribution. It doesn't matter what the probability distributions of the individual stochastic variables is – the sum of enough of them will have a Gaussian distribution.

"The sum of a large number of independent statistical variables" describes thermal noise pretty well. Thermal noise is the sum of a lot of very small random fluctuations in the movement of

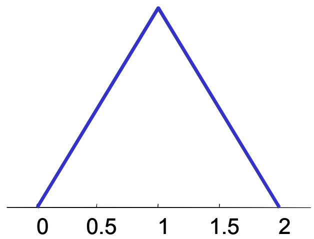
[1] Sometimes called IID variables – IID standing for 'independent identically distributed'.

electrons, and there are so many electrons around that the sum of all of their effects becomes, to a very close approximation, a Gaussian distribution.

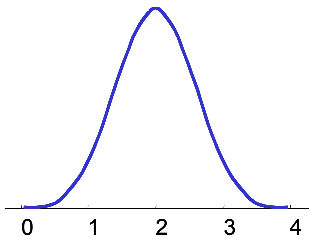
For example: consider a set of random numbers, taken from a uniform distribution from zero to one. Plot the probability distribution of one, then of the sum of two of them, then of the sum of four, and finally of the sum of one hundred such random variables[2]:



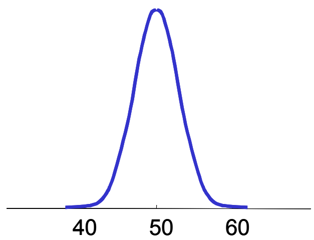
One Uniform Random Variable



Sum of Two Random Variables



Sum of Four Random Variables



Sum of 100 Random Variables

The Gaussian (normal) Distribution	
Probability Distribution:	$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$
Cumulative Probability Distribution:	$P_x(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x-a}{\sqrt{2}\,\sigma}\right)\right)$ $P_x(x) = 1 - \mathrm{Q}\left(\frac{x-a}{\sigma}\right)$
Mean: a	Mode: a
Median: a	Standard Deviation: σ
Variance: σ^2	rms: $\sqrt{\sigma^2 + a^2}$

---

[2] For the interest of MATLAB users: you can generate approximations to these plots using:

```
a=rand(1,1000000);[y,x]=hist(sum(a,1),50);plot(x,y/max(y));axis([-0.25 1.25 0 1.2]);
a=rand(2,500000);[y,x]=hist(sum(a,1),50);plot(x,y/max(y));axis([-0.25 2.25 0 1.2]);
a=rand(4,125000);[y,x]=hist(sum(a,1),50);plot(x,y/max(y));axis([-0.25 4.25 0 1.2]);
a=rand(100,100000);[y,x]=hist(sum(a,1),50);plot(x,y/max(y));axis([30.25 70.25 0 1.2]);
```

The cumulative probability distribution is commonly quoted in terms of two derived functions: the error function erf(x) and the Q-function[3].

(Before we can get much further with this discussion, you'll need to know that:

$$\int_{-\infty}^{\infty} \exp\left(-t^2\right) dt = \sqrt{\pi}$$

which is quite simple to prove if you know the trick[4].)

1.2.1 The erf function

First consider the error function (erf). This is defined as:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-t^2\right) dt$$

---

[3] Why do we need two derived functions to represent the cumulative distribution function? I don't know. Sometimes one is used, sometimes the other. It's a frequent source of confusion. I'll stick to the Q-function here, since it seems more logical to me. MATLAB users beware: the *erf* function is built in, the Q-function is not.

As if that wasn't confusing enough, there are two versions of the Q-function in the literature. I'll use the one most commonly found in engineering textbooks, but be careful when reading some mathematics books, they might be using the other one.

[4] This is quite straightforward to prove if you know the trick: consider a two-dimensional integral of the function:

$$\int_{-\infty}^{\infty} \exp\left(-x^2\right) dx \int_{-\infty}^{\infty} \exp\left(-y^2\right)\ dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-x^2\right) \exp\left(-y^2\right)\ dx\, dy$$

now the "trick" is to change variables to polar co-ordinates, so that $r^2 = x^2 + y^2$, and the elemental area $dx\, dy$ is now $r\, dr\, d\theta$.

$$= \int_0^{\infty} \int_{-\pi}^{\pi} \exp\left(-r^2\right) r\, dr\, d\theta = \int_0^{\infty} \exp\left(-r^2\right) r\, dr \int_{-\pi}^{\pi}\ d\theta$$

$$= 2\pi \int_0^{\infty} \exp\left(-r^2\right) r\, dr = 2\pi \left[\frac{-1}{2} \exp\left(-r^2\right) \right]_0^{\infty} = \pi$$

but since the two integrals we started with must have the same value, they must each equal $\sqrt{\pi}$. Alternatively, note that the integral over all of the normal distribution must be one, so with a standard deviation of $\sqrt{2}$ and a mean of zero, the result follows immediately – however that assumes you trust me about the equation for the distribution.

This can be thought of as twice the value of the integral from zero to some value x of a normal distribution with zero mean, and a standard deviation of $\sqrt{1/2}$. Why twice? Well, that means that erf(x) is the probability that any particular value lies within $x\sqrt{2}$ of the mean of the standard deviation, in either direction. For example: $\operatorname{erf}\left(\sqrt{1/2}\right) = 0.6827...$ is the probability that any variable lies within one standard deviation of the mean.

The CDF of a Gaussian distribution can then be derived as:

$$P_x(X) = \int_{-\infty}^{X} p_x(x)\,dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{X} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)dx$$

and with a change of variable to $t = \frac{(x-a)}{\sqrt{2}\,\sigma}$, we get:

$$P_x(X) = \int_{-\infty}^{X} p_x(x)\,dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\frac{X-a}{\sqrt{2}\sigma}} \exp\left(-t^2\right)\left(\sqrt{2}\sigma dt\right)$$

$$= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\frac{X-a}{\sqrt{2}\sigma}} \exp\left(-t^2\right)dt = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{0} \exp\left(-t^2\right)dt + \frac{1}{\sqrt{\pi}}\int_{0}^{\frac{X-a}{\sqrt{2}\sigma}} \exp\left(-t^2\right)dt$$

$$= \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{X-a}{\sqrt{2}\,\sigma}\right) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{X-a}{\sqrt{2}\,\sigma}\right)\right)$$

which isn't a particularly nice expression.

If you find the choice of a standard deviation of $\sqrt{1/2}$ confusing, and you'd prefer a simpler form of the CDF, then like me you might prefer to use the Q-function.

1.2.2 The Q function

The Q-function is defined as[5]:

$$\mathrm{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_{x}^{\infty} \exp\left(-\frac{t^2}{2}\right)dt$$

which is the integral from some value x to infinity of a Gaussian distribution with a standard deviation of one. That makes the CDF much easier to describe in terms of the Q-function: it's just

---

[5] There's another version sometimes used in mathematics textbooks: $\mathrm{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_{0}^{x} \exp\left(-\frac{t^2}{2}\right)dt$. The difference is a constant value of one-half. I won't be using this one here.

$$P_x(X) = 1 - \mathrm{Q}\left(\frac{X-a}{\sigma}\right)$$

In other words, the Q-function Q(x) is the probability that the value of randomly chosen sample from the a Gaussian distribution with a standard deviation of one has a value of greater than x.

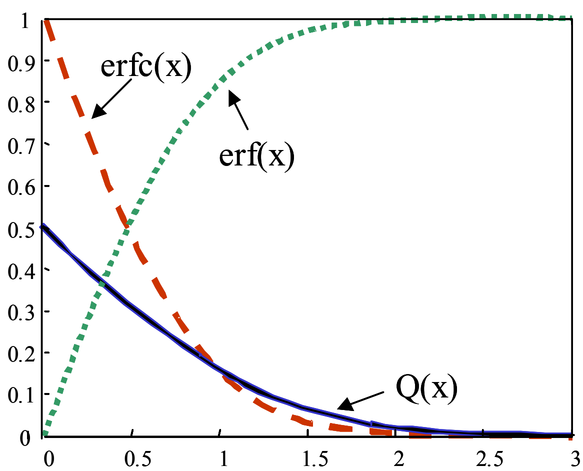
We can expressed the CDF of any Gaussian distribution in terms of the Q-function by using a change of variable in the expression for the CDF of $t = \frac{(x-a)}{\sigma}$, giving:

$$P_x(X) = \int_{-\infty}^{X} p_x(x)\,dx = 1 - \int_{x}^{\infty} p_x(x)\,dx = 1 - \frac{1}{\sqrt{2\pi}\sigma}\int_{\frac{X-a}{\sigma}}^{\infty} \exp\left(-\frac{t^2}{2}\right)(\sigma dt)$$

$$= 1 - \frac{1}{\sqrt{2\pi}}\int_{\frac{X-a}{\sigma}}^{\infty} \exp\left(-\frac{t^2}{2}\right)dt = 1 - \mathrm{Q}\left(\frac{X-a}{\sigma}\right)$$

Hence, it can be readily shown that[6]:

$$\mathrm{Q}(X) = \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{X}{\sqrt{2}}\right)\right) \qquad \mathrm{erf}(X) = 1 - 2\mathrm{Q}\left(X\sqrt{2}\right)$$



1.2.3 *The Log-Normal Distribution*

The log-normal distribution results if the logarithm of the random variable has a normal distribution. It's commonly used in radio-propagation to represent shadowing losses (the loss due to the buildings between the radio transmitter and a random point somewhere around this

---

[6] So MATLAB users can implement the Q-function by using `0.5*erfc(x/sqrt(2));`

(`erfc(x)` is defined as `1 - erf(x)`, and is a MATLAB built-in function as well).

transmitter). It can be derived by considering the probability distribution of a variable *y*, whose logarithm *x* has a normal (Gaussian) distribution. Let $x = \log_e(y)$, then $dx/dy = 1/y$, and:

$$p_y(y) = p_x(x)\frac{dx}{dy} = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)\frac{1}{y}$$

$$p_y(y) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{y}\exp\left(-\frac{(\log_e(y)-a)^2}{2\sigma^2}\right)$$

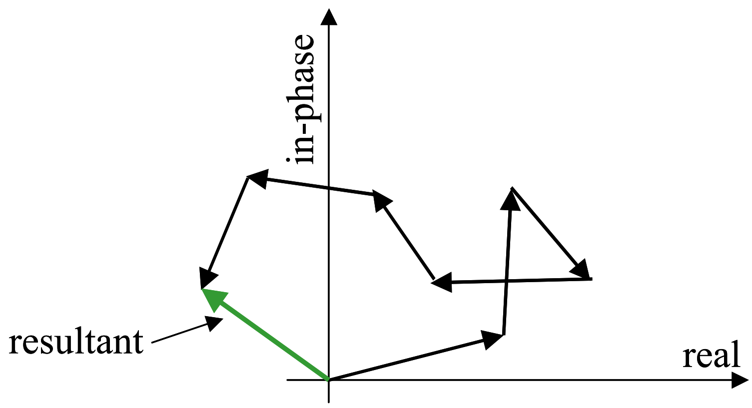
but note here that σ is the standard deviation of the natural logarithm of the variable, and a is the mean of the natural logarithm of the variable. The 'real' standard deviation and mean are tedious to work out, but evaluate to $\exp\left(a+\sigma^2/2\right)\sqrt{\exp\left(\sigma^2\right)-1}$ and $\exp\left(a+\sigma^2/2\right)$ respectively.

The Log-Normal Distribution	
Probability Distribution: $p_y(y) = \begin{cases} 0 & y<0 \\ p_y(y) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{y}\exp\left(-\frac{(\log_e(y)-a)^2}{2\sigma^2}\right) & y \ge 0 \end{cases}$	
Cumulative Probability Distribution: $P_y(Y) = \frac{1}{2}\left(1+erf\left(\frac{\log_e(Y)-a}{\sqrt{2}\,\sigma}\right)\right) \quad Y>0$ $P_x(X) = \frac{1}{2}+Q\left(\frac{\log_e(Y)-a}{\sigma}\right) \quad Y>0$	
Mean: $\exp\left(a+\sigma^2/2\right)$	Mode: $\exp\left(a-\sigma^2\right)$
Median: $\exp(a)$	Standard Deviation: $\exp\left(a+\sigma^2/2\right)\sqrt{\exp\left(\sigma^2\right)-1}$
rms: $\exp\left(a+\sigma^2\right)$	Variance: $\exp\left(2a+\sigma^2\right)\left(\exp\left(\sigma^2\right)-1\right)$

1.3 The Rayleigh Distribution

The Rayleigh distribution is the probability distribution of the amplitude of a complex number, if both the real and imaginary components of the complex number have a normal (Gaussian) distribution with zero mean and the same standard deviation. It's therefore also the probability distribution of the amplitude of the result from adding together a very large number of independent two-dimensional vectors in random directions.

Suppose you have a set of such vectors with completely random directions, all of a similar size. (These could, for example, represent the phases of all the component rays arriving from one radio transmitter and adding together at the receiver). The sum of these vectors is then the received amplitude of the electric field vector at the receiver site:



Consider the component of each of the component vectors along the in-phase axis. These are scalar quantities, with random amplitude and zero mean. By the central limit theorem, the sum of a large number of these rays will tend to be a Gaussian distribution with zero mean. Similarly, the component of the resultant along the quadrature axis will be a Gaussian statistical distribution with zero mean.

Let x be one Gaussian-distributed quantity, and let y be the other. The amplitude of the resultant is therefore $\sqrt{x^2+y^2}$ by Pythagorus.

The probability density of x and y are given by:

$$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-x^2/2\sigma^2\right) \qquad p_y(y) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-y^2/2\sigma^2\right)$$

and since these statistical variables are completely independent, the combined probability density $p_{x,y}(x,y)$ is just:

$$p_{x,y}(x,y) = p_x(x)\,p_y(y) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-(x^2+y^2)/2\sigma^2\right)$$

But $(x^2+y^2) = r^2$, where r is the distance from the origin to the point (x, y), so we can write:

$$p_{x,y}(x,y) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-r^2/2\sigma^2\right)$$

Now to get the total probability that the resultant is a distance r from the origin, we just need to integrate $p_{x,y}(x,y)$ over all points this distance from the origin. Since $p_r(r)\delta r$ is the probability of the resultant having an amplitude of between r and $r+\delta r$, the area over which we have to integrate this function is a thin ring, of width δr, a distance r from the origin. The area of this ring is $2\pi r.\delta r$, hence the answer we require is:

$$p_r(r)\,dr = \frac{2\pi r\,dr}{2\pi\,\sigma^2} \cdot \exp\left(-r^2/2\sigma^2\right)$$

$$p_r(r) = \frac{r}{\sigma^2} \cdot \exp\left(-r^2/2\sigma^2\right)$$

The final point is to notice that the distribution is usually written in terms of the mean power (i.e. the second moment, or the square of the rms value), χ.

The Rayleigh Distribution	
Probability Distribution: $p_r(r) = \begin{cases} 0 & t < 0 \\ \frac{2r}{\chi}\exp\left(-\frac{r^2}{\chi}\right) & t \geq 0 \end{cases}$	
Cumulative Probability Distribution: $P_x(t) = \begin{cases} 0 & t < 0 \\ 1-\exp\left(-\frac{r^2}{\chi}\right) & t \geq 0 \end{cases}$	
Mean: $\sqrt{\frac{\chi\pi}{4}}$	Mode: $\sqrt{\frac{\chi}{2}}$
Median: $\sqrt{\chi \log_e(2)}$	Standard Deviation: $\sqrt{\chi\left(1-\frac{\pi}{4}\right)}$
Variance: $\chi\left(1-\frac{\pi}{4}\right)$	rms: $\sqrt{\chi}$

1.4 The Geometric Distribution

Suppose I have to roll a six before I can start moving my counter around the board in a board game. The number of times I have to throw the die before I get a six is clearly a discrete statistical variable. If the probability of throwing a six is constant, then this variable has a geometric distribution.

In general, suppose the probability of success of some operation is a, where a is constant. Let the probability distribution of the number of attempts required before a success be $q_x(x)$. The probability of succeeding on the first attempt is just a, so $q_x(1) = a$. The probability of succeeding on the second attempt is the probability of failing on the first attempt $(1-a)$ times the probability of succeeding on the second attempt (a)[7], so that $q_x(2) = a(1-a)$. In general, the probability of succeeding on the x^{th} attempt is:

$$q_x(x) = a(1-a)^{x-1}$$

and this is the form of the geometric distribution.

The geometric distribution describes a *memoryless* process, in the sense that each throw of the die is independent, so the system (the die) has no memory of what happened before. This has the initially counter-intuitive result that if you've just thrown a 'one' three times in a row, then the mean number of times you will have to throw the die again before another 'one' turns up is six ($a = 1/6$ in this case, and mean is $1/a$). However, if you haven't thrown a 'one' for the last ten attempts, the mean number of times you have to throw the die before a 'one' turns up is

[7] This is true since the two attempts are identical and independent – the chances of success in the two attempts are identical, and the chance of success in the second attempt does not depend on the result of the first attempt. If this wasn't true, we wouldn't get a geometric distribution.

still six. It doesn't matter what's just happened: the mean number of times you have to wait before getting a 'one' is always six (provided the die is unbiased)[8].

The Geometric Distribution	
Probability Distribution	$q_x(x) = a(1-a)^{x-1} \quad x \geq 1$
Cumulative Probability Distribution	$Q_x(x) = 1-(1-a)^x \quad x \geq 1$
Mean: $\frac{1}{a}$	Standard Deviation: $\frac{\sqrt{(1-a)}}{a}$
Variance: $\frac{(1-a)}{a^2}$	rms: $\frac{\sqrt{2-a}}{a}$

1.4.1 The Negative Exponential Distribution[9]

This distribution can be considered as a limiting case of the geometric distribution in the case of a very large number of trials, each with a very small chance of happening. Suppose the probability of a success in each trial is $\lambda\, dt$, and there is one trial every dt seconds. How long will it be before the next success? The geometric distribution would suggest the distribution of waiting times would be:

$$q_t(t) = \lambda\, dt\,(1-\lambda\, dt)^{\frac{t}{dt}-1} \qquad \text{where } t \gg dt$$

since there are $t\,/\,dt$ trials in every time period t. As dt tends towards zero, this can be approximated using the identity[10]:

---

[8] This runs counter to the 'law of averages' that suggests that, for example, 'lightning never strikes twice in the same place'. The law is entirely wrong.

[9] Sometimes just called the "exponential distribution", since a real exponential function heads off to infinity very rapidly, so isn't much use as a distribution - of positive numbers, anyway.

[10] For those who know what a Maclaurin series is… consider expanding the term as a Maclaurin series (with a non zero value for x):

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \ldots$$

$$\begin{aligned}(1-Nx)^{1/x} &= 1 - \frac{Nx}{1!}\left(\frac{1}{x}\right) + \frac{(Nx)^2}{2!}\left(\frac{1}{x}\right)\left(\frac{1}{x}-1\right) - \frac{(Nx)^3}{3!}\left(\frac{1}{x}\right)\left(\frac{1}{x}-1\right)\left(\frac{1}{x}-2\right) + \ldots \\ &= 1 - N + \frac{N^2}{2!}(1-x) - \frac{N^3}{3!}(1-x)(1-2x) + \ldots\end{aligned}$$

which as x tends to zero, tends towards: $1 - N + \frac{N^2}{2!} - \frac{N^3}{3!} + \ldots$ which is just the Taylor series of *exp(-N)*.

$$\lim_{x \to 0}\left(1 - Nx\right)^{1/x} = \exp\left(-N\right)$$

to give[11]:

$$q_t\left(t\right) = \lambda \ \exp(-\lambda t)\, dt$$

This is the probability that the first success happens in some small time *dt* after waiting for a time *t*. That's the definition of $p_t(t).dt$, where $p_t(t)$ is a continuous probability distribution (remember $p_t(t).dt$ is defined as the probability of the variable taking a value between *t* and *t*+*dt*). So we can write:

$$p_t\left(t\right) dt = \lambda \ \exp(-\lambda t)\, dt$$

$$p_t\left(t\right) = \lambda \ \exp(-\lambda t)$$

Just like the geometric distribution, it doesn't matter when you start waiting for a successful event to happen, the length of time you have to wait is the same. If buses come, entirely independent of each other, on average once an hour, then the average time you will have to wait is one hour.

It doesn't matter if you've been waiting at the bus stop for three days and no bus has turned up, the average length of more time you'll have to wait is one hour. If you've just missed a bus, so you go and have a cup of coffee then come back to the bus stop ten minutes later, the average length of time you'll have to wait at the bus stop is one hour. If you didn't go and have the coffee, the average time you'll have to wait is still one hour. This all results from the fact that the probability that a bus turns up in any small interval of time is constant: it doesn't depend on what's just happened. The system has no "memory".

The Negative Exponential Distribution	
Probability Distribution	$p_t\left(t\right) = \begin{cases} 0 & t < 0 \\ \lambda \ \exp(-\lambda t) & t \ge 0 \end{cases}$
Cumulative Probability Distribution	$P_x(t) = \begin{cases} 0 & t < 0 \\ 1 - \exp\left(-\lambda t\right) & t \ge 0 \end{cases}$
Mean: $\frac{1}{\lambda}$	Mode: Zero
Median: $\frac{\log_e(2)}{\lambda}$	Standard Deviation: $\frac{1}{\lambda}$
Variance: $\frac{1}{\lambda^2}$	rms: $\frac{\sqrt{2}}{\lambda}$

The negative exponential distribution is often used for the length of time between independent events: for example packets arriving at routers and telephone calls being made to an exchange.

[11] Note I've neglected a term in $(1 - \lambda\, dt)^{-1}$ here, since this will tend towards one as *dt* tends towards zero.

It's also used for the length of telephone calls, the assumption being that the probability of a telephone call ending in the next small interval of time is constant, no matter how long the telephone call has lasted up to then.

1.5 The Poisson Distribution

The Poisson distribution occurs in exactly the same physical situations as the negative exponential distribution: a series of totally independent random events which have a constant chance of happening in any small time interval. In these cases the time between events has a negative exponential distribution, and the number of events that occur in a fixed time has a Poisson distribution. It can be derived starting from the same starting point: that the probability of any event occurring in a small time dt is $(a.dt)$.

Then, the probability that there are x successes in a period T is the probability that of the T/dt small time intervals in T, x trials succeed, and $(T/dt - x)$ trials fail. Since they can be any x trials, the probability of this happening is[12]:

$$q_x(x) = \frac{(T/dt)!}{(T/dt - x)!\,x!}(a\,dt)^x\,(1 - a\,dt)^{\frac{T}{dt}-x}$$

and all we have to do to derive the form of the Poisson distribution is to determine the form of this distribution as dt tends to zero. As before, we can use the expression:

$$\lim_{x\to 0}(1 - Nx)^{1/x} = \exp(-N)$$

to simplify the last term; and we can consider the term:

$$\frac{(T/dt)!}{(T/dt - x)!} = \left(\frac{T}{dt}\right)\left(\frac{T}{dt}-1\right)\left(\frac{T}{dt}-2\right)\ldots\left(\frac{T}{dt}-x+1\right)$$

as dt tends to zero will tend towards $(T/dt)^x$, since (T/dt) will become much greater than the number of successes. That gives as dt tends to zero:

[12] If you're not familiar with combinations: this expression is the probability that some particular, defined set of x of the T/dt trials succeed while all other don't (which is $(a\,dt)^x\,(1 - a\,dt)^{\frac{T}{dt}-x}$) times the number of different possible sets of x successes. The first success can be any of the T/dt, the second can be any of the other $(T/dt-1)$, the third any of the remaining $(T/dt-2)$, etc, so that gives $(T/dt)! \,/\, (T/dt-x)!$. However, that would count only one particular order of the successes, and these successes can be in any order: having the first success in the 15th attempt, and the second success in the 29th attempt gives exactly the same pattern of successes as having the first success in the 29th attempt and the second success in the 15th attempt. For x successes, there are $x!$ different ways of ordering the successes: the first can be any of the x, the second any of the remaining $(x-1)$, etc, and we need to divide by this so we only count each pattern of successes once.

$$q_x(x) = \frac{(T/dt)^x}{x!}(a\,dt)^x \exp(-aT + a\,dt\,x)$$

$$q_x(x) = \frac{(aT)^x}{x!}\exp(-aT)$$

which is more usually written in terms of $\lambda = aT$:

The Poisson Distribution	
Probability Distribution: $q_x(x) = \frac{(\lambda)^x}{x!}\exp(-\lambda)$	
Mean: λ	Standard Deviation: $\sqrt{\lambda}$
Variance: λ	rms: $\sqrt{\lambda(1+\lambda)}$

(There isn't a neat closed-form expression for the Cumulative Probability Distribution of a Poisson distribution.)

1.6 Problems

1) Prove that the variance of a uniform distribution with minimum value *a* and maximum value *b* is $\frac{(b-a)^2}{12}$.

2) What is the probability that a sample from a Gaussian distribution with a mean value of 5 and a standard deviation of 3 is above 11? Express your answer in terms of both the Q-function and erf functions.

3) The average intelligence quotient (IQ) of people is 100, with a standard deviation of 15, and a Gaussian distribution. An organisation called MENSA is open to the top 2% of the population. What is the minimum IQ needed to join MENSA?

4) At a school, every pupil in a class of 30 students was asked to run 100 meters. The time it took the students fitted a Gaussian distribution with a mean of 19 seconds and a standard deviation of 3 seconds. There are one million such students in the country – approximately how many of them can run 100 meters in less than five seconds?

5) Derive the formulas given above for the mean, mode, median and variance of the Rayleigh distribution.

6) When my telephone rings, I pick up the phone. During any given second in the subsequent phonecall, the chances of my putting the phone back down are 0.01. What percentage of my phonecalls last more than ten minutes?

7) Every week, I buy one lottery ticket, and I have done this all my life. If the chance of winning a prize in any month is 1/24000, what is the probability that I'll win three prizes in one 70-year lifetime?

8) Consider the Pareto distribution: $p(x) = b^a a / x^{a+1}$, where $b < x < \infty$ (the distribution has a value of zero when x is less than b). Sketch the distribution, and determine the mean and variance as a function of a and b.

Notice anything unusual? For what values of a and b does this distribution have an infinite variance, and an infinite mean? Is this just a mathematical curiosity, or is it possible a distribution with an infinite variance or mean might be useful in engineering?