Distributions

Recall that an integrable function $f : R \rightarrow [0,1]$ such that $\int_R f(x)dx = 1$ is called a ***probability density function*** (pdf). The ***distribution function*** for the pdf is given by

$$F(x) = \int_{-\infty}^{x} f(z)dz$$

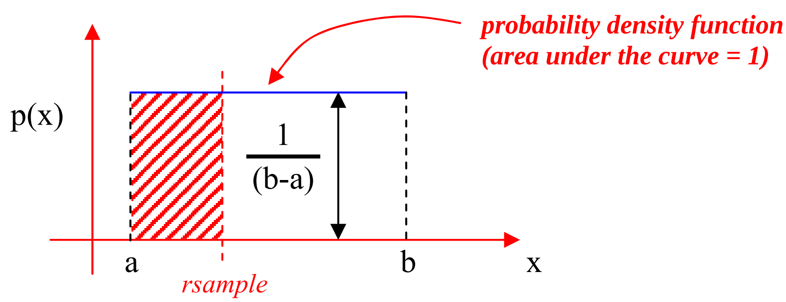
(corresponding to the cumulative distribution function for the discrete case).

Sampling from the distribution corresponds to solving the equation

$$x = \int_{-\infty}^{rsample} f(z)dz$$

for *rsample* given random probability values $0 \le x \le 1$.

I. Uniform Distribution



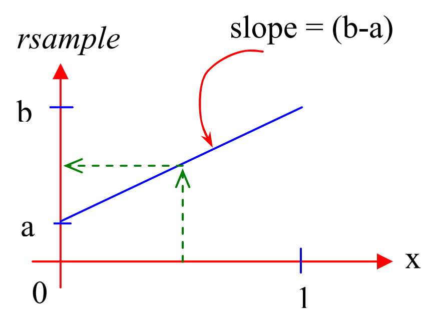
The pdf for values uniformly distributed across [a,b] is given by $f(x) = \dfrac{1}{(b\text{-}a)}$

Sampling from the Uniform distribution:

(pseudo)random numbers x drawn from [0,1] distribute uniformly across the unit interval, so it is evident that the corresponding values

rsample = a + x(b-a)

will distribute uniformly across [a,b].



Note that directly solving

$$x = \int_{-\infty}^{rsample} f(z)dz$$

for *rsample* as per

$$x = \int_{a}^{rsample} \frac{1}{b-a}dz = \left.\frac{z}{b-a}\right]_{a}^{rsample} = \frac{rsample}{b-a} - \frac{a}{b-a}$$

also yields *rsample* = a + x(b-a) (exactly as it should!).

The mean of the uniform distribution is given by

$$\mu = E(X) = \int_a^b z\left(\frac{1}{b-a}\right)dz = \frac{b^2-a^2}{2}\frac{1}{b-a} = \frac{b+a}{2} \quad \textit{(midpoint of [a, b])}$$

z f(z) dz

The standard deviation of the uniform distribution is given by

$$\sigma^2 = E((X-\mu)^2) = \int_a^b \left(z - \frac{b+a}{2}\right)^2 \left(\frac{1}{b-a}\right)dz = \frac{(b-a)^2}{12} \quad \textit{(with some work!)}$$

$(z-\mu)^2$ f(z) dz

II. Normal Distribution

For a finite population the mean (m) and standard deviation (s) provide a measure of average value and degree of variation from the average value. If random samples of size n are drawn from the population, then it can be shown (the Central Limit Theorem) that the distribution of the sample means approximates that of a distribution with

mean: $\mu = m$

standard deviation: $\sigma = \frac{s}{\sqrt{n}}$

pdf:

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

which is called the Normal Distribution. The pdf is characterized by its "bell-shaped" curve, typical of phenomena that distribute symmetrically around the mean value in decreasing numbers as one moves away from the mean. The "empirical rule" is that

- approximately 68% are in the interval $[\mu-\sigma,\mu+\sigma]$
- approximately 95% are in the interval $[\mu-2\sigma,\mu+2\sigma]$
- almost all are in the interval $[\mu-3\sigma,\mu+3\sigma]$

This says that if n is large enough, then a sample mean for the population is accurate with a high degree of confidence, since σ decreases with n. What constitutes "large enough" is largely a function of the underlying population distribution. The theorem assumes that the samples of size n which are used to produce sample means are drawn in a random fashion.

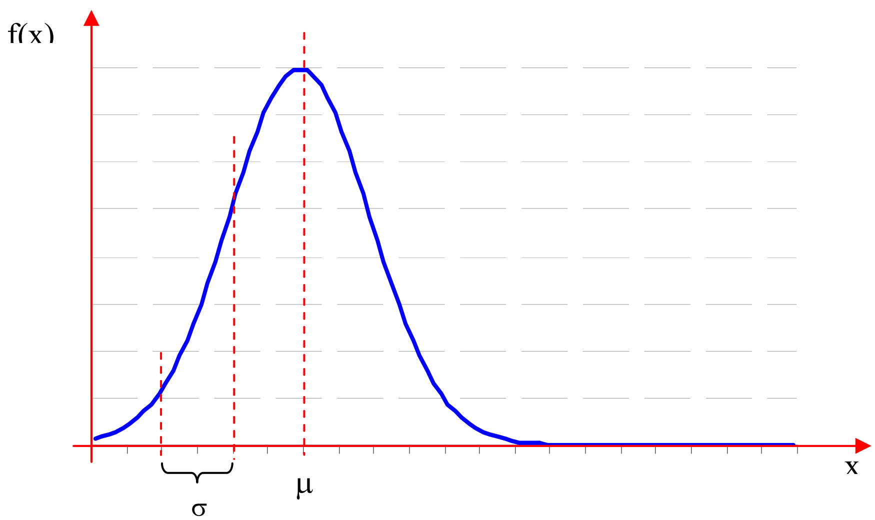
Measurements based on an underlying random phenomena tend to distribute normally. Hogg and Craig (*Introduction to Mathematical Statistics*) note that the kinds of phenomena that have been found to distribute normally include such disparate phenomena as 1) the diameter of the hole made by a drill press, 2) the score on a test, 3) the yield of grain on a plot of ground, 4) the length of a newborn child.

The assumption that grades on a test distribute normally is the basis for so-called "curving" of grades (note that this assumes some underlying random phenomena controlling the measure given by a test; e.g., genetic selection). The practice could be to assign grades of A,B,C,D,F based on how many "standard deviations" separates a percentile score from the mean. Hence, if the mean score is 77.5, and the standard deviation is 8, then the curve of the class scores would be given by

A:	94 and up	(2.5%)
B:	86-93	(13.5%)
C:	70-85	(68%)
D:	62-69	(13.5%)
F:	otherwise	(2.5%)

Most people "pass", but A's are hard to get! This could be pretty distressing if the mean is 95 and the standard deviation is 2 (i.e., 90 is an F).

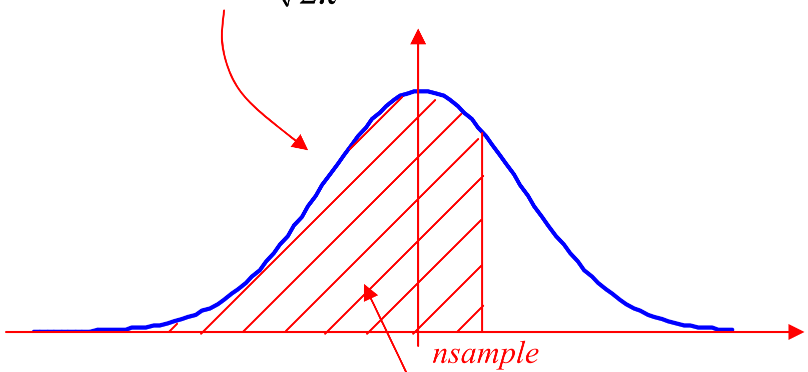
A plot of the pdf for the normal distribution with $\mu = 30$ and $\sigma = 10$ has the appearance:



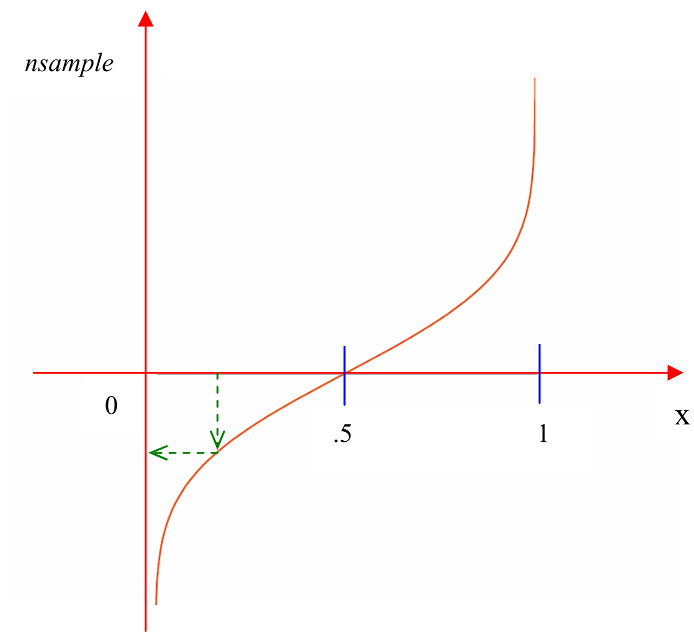
Note that the distribution is completely determined by knowing the value of μ and σ.

The standard normal distribution is given by μ = 0 and σ = 1, in which case the pdf becomes

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



$$\mathbf{x} = \int_{-\infty}^{nsample} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$



It is sufficient to sample from the standard normal distribution, since the linear relationship

$$rsample = \mu + \sigma \cdot nsample$$

holds.

substitute z=μ+σ·t
dz = σdt

$$\int_{-\infty}^{rsample} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(z-\mu)^2}{2}} dz = \int_{-\infty}^{\mu+\sigma \cdot nsample} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{(\mu+\sigma t-\mu)^2}{2\sigma^2}} \sigma dt = \int_{-\infty}^{\mu+\sigma \cdot nsample} \frac{1}{2\pi} \cdot e^{\frac{t^2}{2}} dt$$

There is no "closed-form formula" for *nsample*, so approximation techniques have to be used to get its value.

III. Exponential Distribution

The exponential distribution arises in connection with **Poisson processes**. A Poisson process is one exhibiting a random arrival pattern in the following sense:

1. For a small time interval Δt, the probability of an arrival during Δt is $\lambda\Delta t$, where λ = the mean arrival rate;
2. The probability of more than one arrival during Δt is negligible;
3. Interarrival times are independent of each other.

[this is a kind of "stochastic" process, one for which events occur in a random fashion].

Under these assumptions, it can be shown that the pdf for the distribution of interarrival times is given by

$$f(x) = \lambda e^{-\lambda x}$$

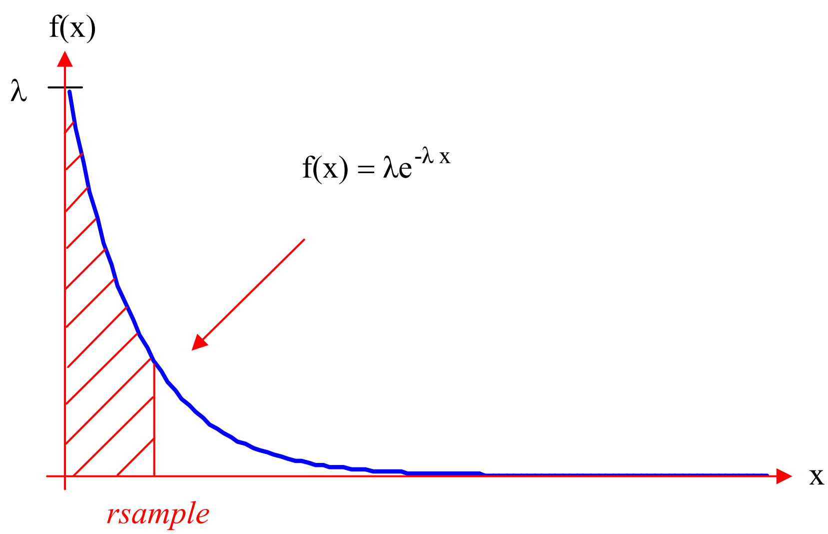
which is the **exponential distribution**.

More to the point, if it can be shown that the number of arrivals during an interval is Poisson distributed (i.e., the arrival *times* are Poisson distributed), then the interarrival *times* are exponentially distributed. Note that the mean arrival rate is given by λ and the mean interarrival time is given by $1/\lambda$. The Poisson distribution is a discrete distribution closely related to the binomial distribution and so will be considered later.

It can be shown for the exponential distribution that the mean is equal to the standard deviation; i.e.,

$$\mu = \sigma = 1/\lambda$$

Moreover, the exponential distribution is the only continuous distribution that is "memoryless", in the sense that $P(X > a+b \mid X > a) = P(X > b)$.



When $\lambda = 1$, the distribution is called the *standard exponential distribution*. In this case, inverting the distribution is straight-forward; e.g.,

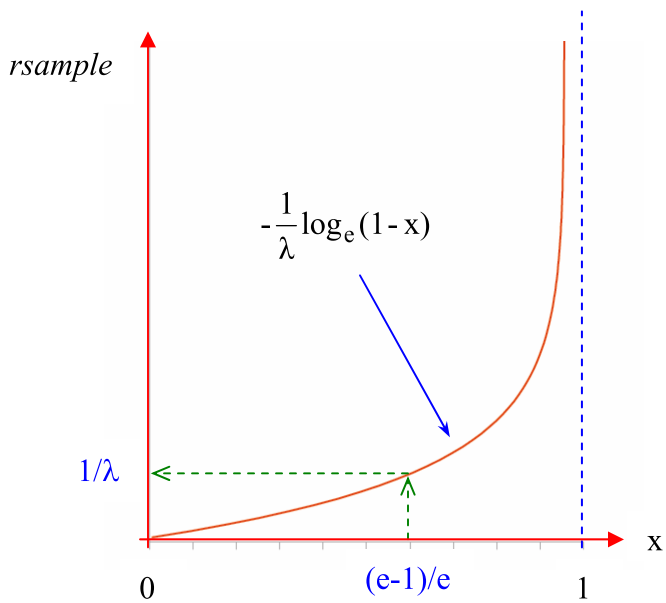
$$\int_0^{nsample} e^{-z}dz = -e^{-z}\Big|_0^{nsample} = 1 - e^{-nsample} = x$$

$$-nsample = \log_e(1-x)$$
$$nsample = -\log_e(1-x)$$

which is a closed form formula for obtaining a normalized sample value (*nsample*) using a random probability x.

General sample values (*rsample*) can then be obtained from the standard exponential distribution by

$$rsample = \frac{1}{\lambda} nsample = -\frac{1}{\lambda}\log_e(1-x)$$



The evident utility of the exponential distribution in discrete systems simulation is its effectiveness for modeling the random arrival pattern represented in a Poisson process. Sampling for interarrival times is a natural approach for introducing new items into the model one at a time. However, care must be taken that when used for this purpose, the exponential distribution is applied to relatively short time periods during which the arrival rate is not dependent on time of day (for example, the model could progress in 1 hour service intervals representing slow, moderate, and peak demand, each governed by an exponential distribution with an appropriate mean interarrival time). A more sophisticated approach is to adjust the arrival rates dynamically with time, a concept studied under the topic of *joint probability distributions*, which will be discussed later.

To verify that **$\mu = \sigma = 1/\lambda$**, integrate by parts to obtain each of μ and σ^2 as follows:

$$\mu = E(X) = \int_0^\infty \underbrace{z}_{u}\, \underbrace{\lambda e^{-\lambda z}\, dz}_{dv} = \underbrace{-ze^{-\lambda z}\Big|_0^\infty}_{uv} + \underbrace{\int_0^\infty e^{-\lambda z} dz}_{-\int vdu} = 0 + \frac{-1}{\lambda} e^{-\lambda x}\Big|_0^\infty = \frac{1}{\lambda}$$

$$\sigma^2 = E\left((X - 1/\lambda)^2\right) = \int_0^\infty \underbrace{(z - 1/\lambda)^2}_{u}\, \underbrace{\lambda e^{-\lambda z} dz}_{dv} = -(z - 1/\lambda)^2\, e^{-\lambda z}\Big|_0^\infty - \int_0^\infty -e^{-\lambda z}\, 2(z - 1/\lambda) dz$$

$$= \frac{1}{\lambda^2} + \frac{2}{\lambda} \underbrace{\int_0^\infty z\,\lambda\, e^{-\lambda z} dz}_{= 1/\lambda} - \frac{2}{\lambda^2} \underbrace{\int_0^\infty \lambda e^{-\lambda z} dz}_{= 1 \text{ (by definition)}} = \frac{1}{\lambda^2}$$

For the pdf of the exponential distribution $f(x) = \lambda e^{-\lambda x}$ note that $f'(x) = -\lambda^2 e^{-\lambda x}$ so $f(0) = \lambda$ and $f'(0) = -\lambda^2$

Hence, if $\lambda < 1$ the curve starts lower and flatter than for the standard exponential. The asymptotic limit is the x-axis.