R-Based Probability Distributions

General Comments

When a parameter name is followed by an equal sign, the value given is the default.

Consider a random variable that has the range, $a \le x \le b$. The parameter, `lower.tail`, when `TRUE` means that the cumulative distribution function is to be interpreted as running from *a* to some specified value of $x = x'$. When the parameter is `FALSE` the cumulative distribution function is to be interpreted as running from x' to *b*.

Probability density functions all start with the letter *d*. Example, `dbinom()`.
Cumulative distribution functions all start with the letter *p*. Example, `pnorm()`.
Inverse cumulative distribution functions all start with the letter *q* (which is short for quantile).
Example, `qt()`.
Random number generators all start with the letter *r*. Example, `runif()`.
Full examples of function names are given under **Binomial**.

Available Distributions

Enter `?Distributions` to get a complete listing. Common distributions in chemical data processing are:

binomial: `binom()`, occurs when an experiment has two outcomes, e.g. success or failure. It deals with the distribution of successes out of a specified number of trials. The random variable is discrete. An example is solvent extraction.

Cauchy (Lorentzian): `cauchy()`, occurs when observing spectroscopic line shapes, or the Fourier transform of a free-induction decay.

exponential: `exp()`, results from a first order process.

F-statistic: `f()`, occurs when comparing two experimental variances.

normal (Gaussian): `norm()`, the bell-shaped curve that represents the distribution of thermal noise.

Poisson: `pois()`, the distribution of counting data. The random variable is discrete.

t-statistic: `t()`, occurs when comparing data normalized by the experimental standard deviation.

uniform: `unif()`, occurs when a random variable has a constant probability over its entire range, e.g. rolling a die. The random variable can be discrete or continuous.

Binomial

This *discrete* density function is given by,

$$f\left(x,n,p\right)=\frac{n!}{x!\left(n-x\right)!}p^{x}\left(1-p\right)^{n-x}$$

where *x* is the number of successes, *n* is the number of trials (called *size* in R), and *p* is the probability of observing a success. The range of the random variable is $0 \le x \le n$. In R the factorial term is given by the strangely-named function, `choose(n,x)`. The mean is given by *np* and the variance by *np*(1 - *p*).

In the following function calls, the notation ... signifies that there are less commonly used parameters.

`dbinom(x,size,probability,...)`, `x` is the number of successes, `size` is the number of trials, and `probability` is the probability of observing a success. As an example, `dbinom(5,10,0.5)` is the probability of observing 5 heads when tossing a coin 10 times.

`pbinom(q,size,probability)`, the probability of observing `0` to `q` successes out of `size` tries, where `probability` is the probability of observing a success. As an example, `pbinom(5,10,0.5)` is the probability of observing 0 through 5 heads when tossing a coin ten times.

`qbinom(p,size,probability)`, the value of `q` required so that observing `0` to `q` successes has a probability of `p`. Again, `size` is the number of tries and `probability` is the probability of observing a success. As an example, `q <- qbinom(0.5,10,0.5)` yields `q = 5` where the range 0 to 5 successes is observed with a 0.5 probability, when tossing a coin 10 times.

`rbinom(n,size,probability)`, generates `n` random successes ranging from `0` to `size` in value, when `probability` is the probability of a success. As an example, `rbinom(8,10,0.5)` yields, `(6, 8, 5, 4, 4, 5, 5, 6)` which is the number of heads observed for each of 8 replications of an experiment where a coin is tossed ten times.

Cauchy

This *continuous* density function is given by,

$$f\left(x,\mu,\Gamma\right)=\frac{1}{\pi}\frac{\Gamma/2}{\left(\Gamma/2\right)^{2}+\left(x-\mu\right)^{2}}$$

where *x* is the random variable, μ is the mean, Γ is the full-width-at-half-maximum (FWMH or peak width). The mean is μ. The distribution does not have a finite variance.

`dcauchy(x, location=0,scale=1,...)`, `x` is the random variable, `location` is μ in the above equation, and `scale` is $\Gamma/2$.

All other functions follow the form given under **Binomial**.

Exponential

This *continuous* density function has two common forms,

$$f(x,\lambda) = \lambda \exp(-\lambda x) \qquad f(t,\tau) = (1/\tau)\exp(-t/\tau)$$

where *x* or *t* are the random variable, $1/\lambda$ (or τ) is the mean and $1/\lambda^2$ (or τ^2) is the variance.

Chemists usually call λ the rate constant and τ the lifetime, where $\lambda = 1/\tau$.

Note: $0 \le x,t \le \infty$.

`dexp(x,rate=1,...)`, `x` is the random variable and rate is `λ`.

All other functions follow the form given under **Binomial**.

F-Statistic

This *continuous* density has the form,

$$f(x,\phi_1,\phi_2) = \frac{\Gamma\left(\frac{\phi_1+\phi_2}{2}\right)}{\Gamma(\phi_1/2)\Gamma(\phi_2/2)}\left(\frac{\phi_1}{\phi_2}\right)^{\frac{\phi_1}{2}} x^{\frac{\phi_1-2}{2}}\left(1+\left(\frac{\phi_1}{\phi_2}\right)x\right)^{-\frac{\phi_1+\phi_2}{2}}$$

where *x* is the random variable, φ_1 is the numerator degrees of freedom, and φ_2 is the denominator degrees of freedom. Γ () is the gamma function (in R, `gamma()`).

The mean and variance are of little importance. Note: since variances are positive, $0 \le x \le \infty$.

`df(x,df1,df2,...)`, `x` is the random variable, `df1` is the numerator degrees of freedom, and `df2` is the denominator degrees of freedom.

All other functions follow the form given under **Binomial**.

Normal (Gaussian)

This *continuous* density has the form,

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

where *x* is the continuous random variable, μ is the mean, and σ^2 is the variance.

`dnorm(x,mean=0,sd=1,...)` , `x` is the random variable, `mean` is the mean, and `sd` is the standard deviation.

All other functions follow the form given under **Binomial**.

Poisson

This *discrete* Poisson or counting density has the form,

$$f(x,\mu) = \frac{\mu^x}{x!}\exp(-\mu)$$

where *x* is the random variable, *μ* is the mean *and* variance. Note: 0 ≤ *x* ≤ ∞.

`dpois(x,lambda,...)`, `x` is the random variable and `lambda` is the mean.
All other functions follow the form given under **Binomial**.

t-Statistic

This *continuous* density has the form,

$$f(x,\phi) = \frac{1}{\sqrt{\phi\pi}} \frac{\Gamma\left(\frac{\phi+1}{2}\right)}{\Gamma\left(\frac{\phi}{2}\right)} \left(1+\frac{x^2}{\phi}\right)^{-\frac{\phi+1}{2}}$$

where *x* is the random variable and *φ* is the degrees of freedom. The mean and variance are of little importance. *Γ* is the gamma function (not the gamma density!).

`dt(x,df,...)`, `x` is the random variable and `df` is the degrees of freedom.
All other functions follow the form given under **Binomial**.

Uniform

The *discrete* and *uniform* density functions both have the form,

$$f(x,a,b) = \frac{1}{b-a}$$

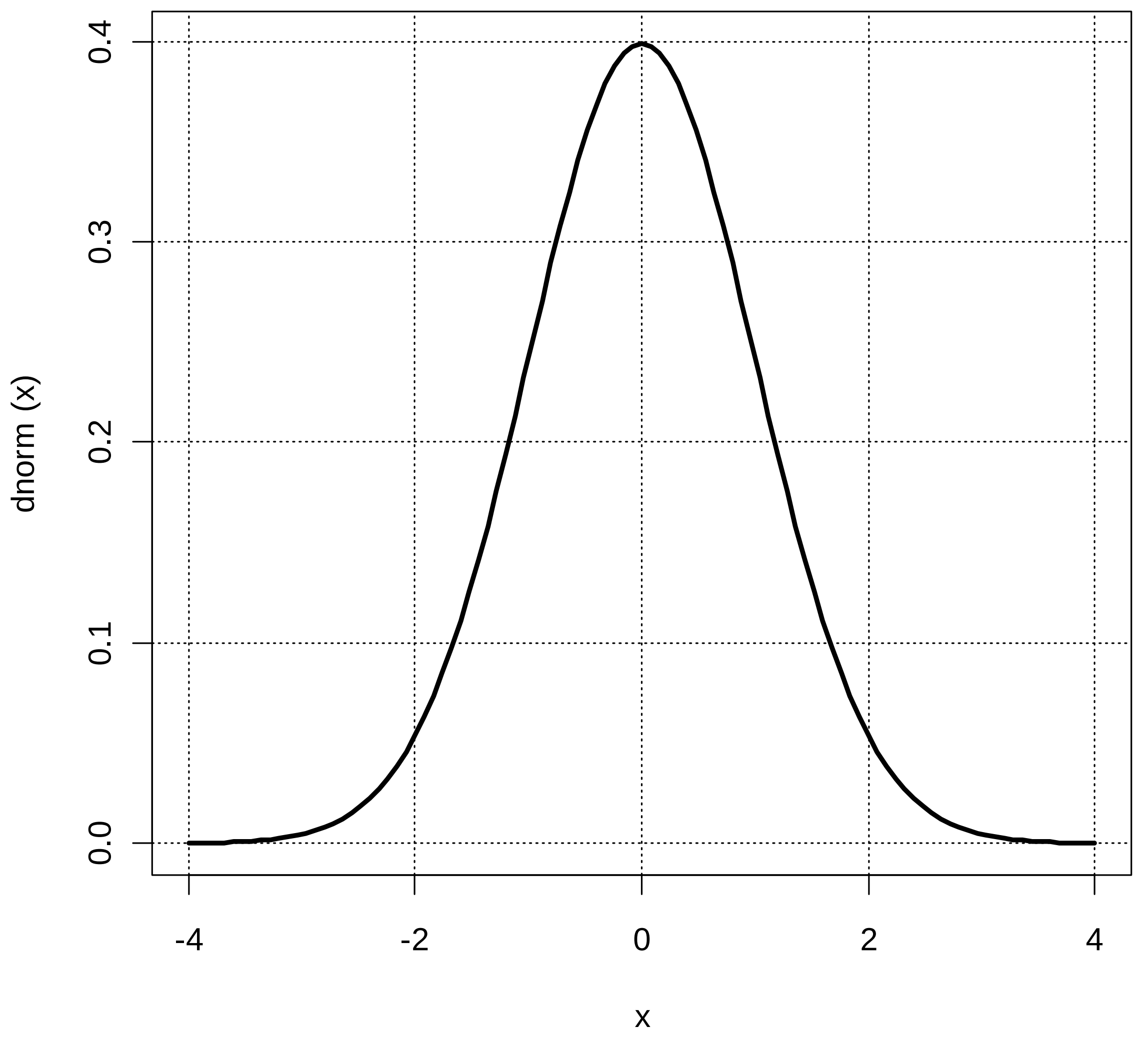
where *x* is the random variable over the range, $a < x \le b$. For a discrete variable it is important that the lower bound, *a*, is not included. This distinction makes no difference with a continuous random variable since < and ≤ are only off by the infinitely small amount, *dx*. The means are $(b + a + 1)/2$ for discrete and $(b + a)/2$ for continuous distributions. The variances are $[(b - a)^2 - 1]/12$ for discrete and $(b - a)^2/12$ for continuous.

`dunif(x,min=0,max=1,...)`, where `x` is the *continuous* random variable, `min` is *a*, and `max` is *b*.
All other functions follow the form given under **Binomial**.

Visualizing the Density Functions

For *continuous* density functions a quick and dirty way to visualize them is to use the `curve()` function. For example,

```
curve(dnorm,-4,4,lwd=3)
grid(col='black')
```



For *discrete* density functions use the `barplot()` function. For example,

```
x <- 0:20
p <- dpois(x,5)
barplot(p,names.arg=as.character(x))
abline(h=c(0.05,0.10,0.15),lty=3,col='black')
```