**Common Distributions**

### By Devan Becker

# Discrete Distributions

|  |  |  |
| --- | --- | --- |
| Binomial Distribution This distribution is used to mode the number of successes in a sequence of trials. is the number of successes out of trials, each trial having a probability of success . Trials are assumed to be independent. | | |
|  |  |  |
|  |  |
|  |  |
|  |  |
| R params | size = , prob = |
| Related Distributions  * If are *Bernoulli* random variables, then | |

|  |  |  |
| --- | --- | --- |
| Geometric Distribution This distribution is used to model the number of failures before a success. is the number of *trials*, where there are failures and the last trial was a success. Each trial has a probability of success . Trials are assumed to be independent. | | |
|  |  |  |
|  |  |
|  |  |
|  |  |
| R params | prob = |
| Alternative Parameterization  * can be the number failures, rather than the number of trials. | |

|  |  |  |
| --- | --- | --- |
| Hypergeometric Distribution If you have a jar with red marbles and green marbles ( balls total) and you take one marble out at a time until you have marbles, then the number of red marbles is modelled by a Hypergeometric distribution. | | |
|  |  |  |
|  |  |
|  |  |
|  | Not useful. |
| R params | m = , n = , k = |
| Related Distributions  * If you take the marbles out *with replacement*, you would use a binomial distribution. | |

|  |  |  |
| --- | --- | --- |
| Negative Binomial Distribution This distribution models the number of successes before the a prespecified number of failures (e.g., the number of times a gambler must lose before they give up). is the number of successes before the th failure, where and each trial has a probability of success of . | | |
|  |  |  |
|  |  |
|  |  |
|  |  |
| R params | size = , prob = |
| Related Distributions  * The moment generating function looks awfully similar to that of the geometric distribution… | |

|  |  |  |
| --- | --- | --- |
| Poisson Distribution This distribution models counts, such as the number of calls in a day or the number of fires per year. is the average count. This is a very simple distribution where the mean and variance are assumed equal. | | |
|  |  |  |
|  |  |
|  |  |
|  |  |
| R params | lambda = |
| Related Distributions  * This is a special case of the negative binomial. * The Poisson distribution approximates the binomial and geometric distributions. * If and , are independent, then | |

# Continuous Distributions

|  |  |  |
| --- | --- | --- |
| Normal Distribution Thanks to the central limit theorem, this distribution models basically everything.  **Notation:** , and . | | |
|  |  |  |
|  | For and |
|  |  |
|  |  |
|  |  |
| R params | mean = sd = |
| Related Distributions  * If , then . * All of them. | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gamma Family of Distributions  * The Gamma distribution arises everywhere in mathematical statistics but isn’t often used to model data. It is parameterised by a shape parameter and a rate parameter , which is sometimes expressed in terms of the scale parameter . * The exponential distribution (blue lines in the plot) is useful for the time between events. * The Chi-Squared distribution (solid lines in the plot, where ) is used for variance parameters. | | | | |
| Notation: | Gamma, | Exp( =  Gamma( | =  Gamma |  |
|  |  |  |  |
|  | No simple form. |  | No simple form. |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| R params | shape = rate =, scale = | rate = | df = |

|  |  |  |
| --- | --- | --- |
| Beta Distribution The beta distribution almost looks like it should be in the Gamma family, but it is not. Again, this comes up a lot in mathematical statistics, but is not often used for data. The support of this distribution is (0,1), so this can be used to model probabilities.  **Notation:** Beta(, where both and are both positive and referred to as shape parameters. | | |
|  |  |  |
|  | No simple form. |
|  |  |
|  |  |
|  | Not useful. |
| R params | shape1 = shape2 = |
| Related Distributions  * If , this is the uniform distribution. * A lot of fun transformations into more complicated distributions. * If is much larger than , this is approximated by a normal distribution. | |

|  |  |  |
| --- | --- | --- |
| Weibull Distribution The Weibull distribution is an extension of the Exponential distribution, where the variance is no longer constrained to be the square of the mean, but we are still modelling the mean parameter (unlike with Gamma).  **Notation:** Weibull(, where is the shape and is the scale parameter. | | |
|  | ; |  |
|  |  |
|  |  |
|  |  |
|  | Not useful. |
| R params | shape = scale = |
| Related Distributions  * If , this is the Exponential distribution. * This is the two-parameter Weibull model, there is also a three-parameter model. | |

|  |  |  |
| --- | --- | --- |
| Lognormal Distribution Confusingly, the Lognormal distribution is not the log of a normal distribution. Instead, taking the log of this distribution results in a normal distribution (this naming convention is used a lot). Also, and are the mean and variance of the log of this distribution. This is often used for survival analysis (like the exponential and Weibull distributions).  **Notation:** Lognormal(, where is the location (NOT mean) and is the shape parameter. | | |
|  | ; |  |
|  | , where is the cdf of a standard normal distribution. |
|  |  |
|  |  |
|  | Not useful. |
| R params | mean = sd = |
| Related Distributions  * If is small, this can be approximated by a normal distribution (which is weird). | |

|  |  |  |
| --- | --- | --- |
| Uniform Distribution I have included this distribution simply because I always forget the variance.  **Notation:** Unif(a,b) | | |
|  | ; | Uniform Distribution  ---------------------------------------------------  | |  | |  | |  a b |
|  |  |
|  |  |
|  |  |
|  |  |
| R params | min = a max = |
| Related Distributions  * If is small, this can be approximated by a normal distribution (which is weird). | |