

# Discrete Probability Distributions in Sage

Caleb Frambach

University of Washington

June 2, 2014

# What Sage Previously Used:

Currently Sage just has a `GeneralDiscreteDistribution` class.

So, you have to put probabilities in an array such as `[0.3, 0.4, 0.01 ...]`.

Then it takes the first element as  $P[0] = .3$ ,  $P[1] = .4$  ...

This could be quite lengthy if your probability space is very large and in general, this isn't the most intuitive thing to do.

# My project

I made it so you can create representations of some common discrete distributions with just a few parameters. This is much quicker than what was previously done.

Also, from these common distributions, I created various methods that could be called, all of which are much quicker than the general route. For example, when finding expectation, you would have to normally sum up each event  $x$  multiplied by it's probability.

But, all of the common distributions have shortcuts to find there expection, and this will save time if large numbers ever have to be calculated.

# Distributions

Binomial (From this Bernouilli) : Takes two parameters,  $n$ , the number of events, and  $p$ , the probability of each event.

Uniform: One parameter  $N$ , that says every event has the same probability of occurring over  $n$ .

Poisson: Takes one parameter  $\lambda$  that tells the expected number of events to occur during a timeframe.

Geometric: Takes one parameter  $p$ , the probability of each event. Will tell the chances of  $X-1$  failures before 1 succes.

Negative Binomial: Takes two parameters  $r$ ,  $p$ , where  $p$  is the probability of each event occurring, and  $r$  is the number of failures until the experiment is stopped.  $X$  will be the number of successes trying to happen before the  $r$  failures.

# More Distributions

Hypergeometric: Takes three parameters  $M, K, n$ , and then tell the probability of  $x$  successes, out of  $K$  possible good outcomes, with  $n$  attempts at a population of size  $N$ .

Logarithmic: Takes only one parameter,  $p$ , the probability the event occurs. Represents events that occur exponentially.

Borel: Takes one parameter,  $\mu$  that is used generally in queueing and branching problems.

# Some Methods

In General:

Expectation:  $\sum x * p(x)$

Variance:  $\sum x^2 * p(x)$

Moment Generating Function:  $\sum e^{tx} * p(x)$

But for all of the above distributions, we can use shortcuts to make this process quicker. We don't have to carry out these summations every time.

# Some Examples

Expectation of Binomial:  $n * p$

Expectation of Hypergeometric:  $n * \frac{K}{M}$

Variance of Poisson:  $\lambda$

MGF of Geometric:  $\frac{pe^t}{1-(1-p)e^t}$

These will obviously be much quicker.

# Calculate Values

Most importantly, we can take any value  $X$ , and calculate the probability of that number of events happening.

There are quick formulas for the probability mass functions. Also can look at the cumulative distribution functions.

Some Examples:

Binomial:  $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$

Uniform:  $P(X = x) = \frac{1}{n}$

Geometric:  $P(X = x) = (1 - p)^{x-1} (p)$



# Real World Usage

**Binomial:** Anything that has success and failure. Could be as simple as flipping a coin, rolling a 5 on a die, or as important as winning the lottery or a disease being cured by a vaccine.

**Poisson:** Can model the chance of  $X$  events occurring during a time frame. Important with designs of streets to control traffic.

**Uniform:** Useful for checking the probability of certain things occur when everything is equal. Especially important when mixed with other distributions.

**Negative Binomial:** Can model the number of days a machine may work until it breaks down. Able to give a realistic probability.

**Hypergeometric:** Anything that is sampling without replacement. Can range from poker, to sampling demographics, to just picking marbles out of an urn.

**Geometric:** Probability of how many failures til one succes.

# Questions?

Does anyone have any questions?