Discrete Statistical Distributions

November 21, 2005

Discrete random variables take on only a countable number of values. The commonly used distributions are included in SciPy and described in this document. Each discrete distribution can take one extra integer parameter: L. The relationship between the general distribution and the standard one is

$$p(x) = p_0(x - L)$$

which allows for shifting of the input. When a distribution generator is initialized, the discrete distribution can either specify the beginning and ending (integer) values a and b which must be such that

$$p_0(x) = 0 \quad x < a \text{ or } x > b$$

in which case, it is assumed that the pdf function is specified on the integers $a + mk \leq b$ where k is a non-negative integer $(0, 1, 2, \ldots)$ and m is a positive integer multiplier. Alternatively, the two lists x_k and $p(x_k)$ can be provided directly in which case a dictionary is set up internally to evaulate probabilities and generate random variates.

0.1 Probability Mass Function (PMF)

The probability mass function of a random variable X is defined as the probability that the random variable takes on a particular value.

$$p(x_k) = P[X = x_k]$$

This is also sometimes called the probability density function, although technically

$$f(x) = \sum_k p(x_k)\,\delta(x - x_k)$$

is the probability density function for a discrete distribution[1].

0.2 Cumulative Distribution Function (CDF)

The cumulative distribution function is

$$F(x) = P[X \leq x] = \sum_{x_k \leq x} p(x_k)$$

and is also useful to be able to compute. Note that

$$F(x_k) - F(x_{k-1}) = p(x_k)$$

0.3 Survival Function

The survival function is just

$$S(x) = 1 - F(x) = P[X > k]$$

the probability that the random variable is strictly larger than k.

[1] Note that we will be using p to represent the probability mass function and a parameter (a probability). The usage should be obvious from context.

0.4 Percent Point Function (Inverse CDF)

The percent point function is the inverse of the cumulative distribution function and is

$$G(q) = F^{-1}(q)$$

for discrete distributions, this must be modified for cases where there is no x_k such that $F(x_k) = q$. In these cases we choose $G(q)$ to be the smallest value $x_k = G(q)$ for which $F(x_k) \geq q$. If $q = 0$ then we define $G(0) = a - 1$. This definition allows random variates to be defined in the same way as with continuous rv's using the inverse cdf on a uniform distribution to generate random variates.

0.5 Inverse survival function

The inverse survival function is the inverse of the survival function

$$Z(\alpha) = S^{-1}(\alpha) = G(1-\alpha)$$

and is thus the smallest non-negative integer k for which $F(k) \geq 1-\alpha$ or the smallest non-negative integer k for which $S(k) \leq \alpha$.

0.6 Hazard functions

If desired, the hazard function and the cumulative hazard function could be defined as

$$h(x_k) = \frac{p(x_k)}{1 - F(x_k)}$$

and

$$H(x) = \sum_{x_k \leq x} h(x_k) = \sum_{x_k \leq x} \frac{F(x_k) - F(x_{k-1})}{1 - F(x_k)}.$$

0.7 Moments

Non-central moments are defined using the PDF

$$\mu'_m = E[X^m] = \sum_k x_k^m p(x_k).$$

Central moments are computed similarly $\mu = \mu'_1$

$$\begin{aligned} \mu_m = E\left[(X-\mu)^2\right] &= \sum_k (x_k - \mu)^m p(x_k) \\ &= \sum_{k=0}^{m} (-1)^{m-k} \left(\begin{array}{c} m \\ k \end{array}\right) \mu^{m-k} \mu'_k \end{aligned}$$

The mean is the first moment

$$\mu = \mu'_1 = E[X] = \sum_k x_k p(x_k)$$

the variance is the second central moment

$$\mu_2 = E\left[(X-\mu)^2\right] = \sum_{x_k} x_k^2 p(x_k) - \mu^2.$$

Skewness is defined as

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

while (Fisher) kurtosis is

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3,$$

so that a normal distribution has a kurtosis of zero.

0.8 Moment generating function

The moment generating funtion is defined as

$$M_X(t) = E\left[e^{Xt}\right] = \sum_{x_k} e^{x_k t} p(x_k)$$

Moments are found as the derivatives of the moment generating function evaluated at 0.

0.9 Fitting data

To fit data to a distribution, maximizing the likelihood function is common. Alternatively, some distributions have well-known minimum variance unbiased estimators. These will be chosen by default, but the likelihood function will always be available for minimizing.

If $f_i(k;\boldsymbol{\theta})$ is the PDF of a random-variable where $\boldsymbol{\theta}$ is a vector of parameters (*e.g.* L and S), then for a collection of N independent samples from this distribution, the joint distribution the random vector $\mathbf{k}$ is

$$f(\mathbf{k};\boldsymbol{\theta}) = \prod_{i=1}^{N} f_i(k_i;\boldsymbol{\theta}).$$

The maximum likelihood estimate of the parameters $\boldsymbol{\theta}$ are the parameters which maximize this function with $\mathbf{x}$ fixed and given by the data:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} f(\mathbf{k};\boldsymbol{\theta}) \\ &= \arg\min_{\boldsymbol{\theta}} l_{\mathbf{k}}(\boldsymbol{\theta}).\end{aligned}$$

Where

$$\begin{aligned}l_{\mathbf{k}}(\boldsymbol{\theta}) &= -\sum_{i=1}^{N} \log f(k_i;\boldsymbol{\theta}) \\ &= -N\overline{\log f(k_i;\boldsymbol{\theta})}\end{aligned}$$

0.10 Standard notation for mean

We will use

$$\overline{y(\mathbf{x})} = \frac{1}{N}\sum_{i=1}^{N} y(x_i)$$

where N should be clear from context.

0.11 Combinations

Note that

$$k! = k\cdot(k-1)\cdot(k-2)\cdot\cdots\cdot 1 = \Gamma(k+1)$$

and has special cases of

$$\begin{aligned}0! &\equiv 1 \\ k! &\equiv 0 \quad k<0\end{aligned}$$

and

$$\left(\begin{array}{c} n \\ k \end{array}\right) = \frac{n!}{(n-k)!k!}.$$

If $n<0$ or $k<0$ or $k>n$ we define $\left(\begin{array}{c} n \\ k \end{array}\right) = 0$

1 Bernoulli

A Bernoulli random variable of parameter p takes one of only two values $X = 0$ or $X = 1$. The probability of success ($X = 1$) is p, and the probability of failure ($X = 0$) is $1 - p$. It can be thought of as a binomial random variable with $n = 1$. The PMF is $p(k) = 0$ for $k \neq 0, 1$ and

$$\begin{aligned}
p(k;p) &= \begin{cases} 1-p & k=0 \\ p & k=1 \end{cases} \\
F(x;p) &= \begin{cases} 0 & x<0 \\ 1-p & 0\leq x<1 \\ 1 & 1\leq x \end{cases} \\
G(q;p) &= \begin{cases} 0 & 0\leq q<1-p \\ 1 & 1-p\leq q\leq 1 \end{cases} \\
\mu &= p \\
\mu_2 &= p(1-p) \\
\gamma_3 &= \frac{1-2p}{\sqrt{p(1-p)}} \\
\gamma_4 &= \frac{1-6p(1-p)}{p(1-p)}
\end{aligned}$$

$$M(t) = 1 - p\left(1 - e^t\right)$$

$$\mu'_m = p$$

$$h[X] = p\log p + (1-p)\log(1-p)$$

2 Binomial

A binomial random variable with parameters (n, p) can be described as the sum of n independent Bernoulli random variables of parameter p;

$$Y = \sum_{i=1}^{n} X_i.$$

Therefore, this random variable counts the number of successes in n independent trials of a random experiment where the probability of success is p.

$$\begin{aligned}
p(k;n,p) &= \binom{n}{k} p^k (1-p)^{n-k} \quad k\in\{0,1,\ldots n\}, \\
F(x;n,p) &= \sum_{k\leq x} \binom{n}{k} p^k (1-p)^{n-k} = I_{1-p}(n-\lfloor x\rfloor, \lfloor x\rfloor+1) \quad x\geq 0
\end{aligned}$$

where the incomplete beta integral is

$$I_x(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1}\,dt.$$

Now

$$\begin{aligned}
\mu &= np \\
\mu_2 &= np\left(1-p\right) \\
\gamma_1 &= \frac{1-2p}{\sqrt{np\left(1-p\right)}} \\
\gamma_2 &= \frac{1-6p\left(1-p\right)}{np\left(1-p\right)}.
\end{aligned}$$

$$M\left(t\right) = \left[1-p\left(1-e^t\right)\right]^n$$

3 Boltzmann (truncated Planck)

$$\begin{aligned}
p\left(k;N,\lambda\right) &= \frac{1-e^{-\lambda}}{1-e^{-\lambda N}}\exp\left(-\lambda k\right) \quad k\in\{0,1,\ldots,N-1\} \\
F\left(x;N,\lambda\right) &= \begin{cases} 0 & x<0 \\ \frac{1-\exp\left[-\lambda\left(\lfloor x\rfloor+1\right)\right]}{1-\exp\left(-\lambda N\right)} & 0\leq x\leq N-1 \\ 1 & x\geq N-1 \end{cases} \\
G\left(q,\lambda\right) &= \left\lceil -\frac{1}{\lambda}\log\left[1-q\left(1-e^{-\lambda N}\right)\right]-1\right\rceil
\end{aligned}$$

Define $z=e^{-\lambda}$

$$\begin{aligned}
\mu &= \frac{z}{1-z}-\frac{Nz^N}{1-z^N} \\
\mu_2 &= \frac{z}{\left(1-z\right)^2}-\frac{N^2z^N}{\left(1-z^N\right)^2} \\
\gamma_1 &= \frac{z\left(1+z\right)\left(\frac{1-z^N}{1-z}\right)^3-N^3z^N\left(1+z^N\right)}{\left[z\left(\frac{1-z^N}{1-z}\right)^2-N^2z^N\right]^{3/2}} \\
\gamma_2 &= \frac{z\left(1+4z+z^2\right)\left(\frac{1-z^N}{1-z}\right)^4-N^4z^N\left(1+4z^N+z^{2N}\right)}{\left[z\left(\frac{1-z^N}{1-z}\right)^2-N^2z^N\right]^2}
\end{aligned}$$

$$M\left(t\right) = \frac{1-e^{N(t-\lambda)}}{1-e^{t-\lambda}}\frac{1-e^{-\lambda}}{1-e^{-\lambda N}}$$

4 Planck (discrete exponential)

Named Planck because of its relationship to the black-body problem he solved.

$$\begin{aligned}
p\left(k;\lambda\right) &= \left(1-e^{-\lambda}\right)e^{-\lambda k} \quad k\lambda\geq 0 \\
F\left(x;\lambda\right) &= 1-e^{-\lambda\left(\lfloor x\rfloor+1\right)} \quad x\lambda\geq 0 \\
G\left(q;\lambda\right) &= \left\lceil -\frac{1}{\lambda}\log\left[1-q\right]-1\right\rceil.
\end{aligned}$$

$$
\begin{aligned}
\mu &= \frac{1}{e^{\lambda}-1} \\
\mu_2 &= \frac{e^{-\lambda}}{\left(1-e^{-\lambda}\right)^2} \\
\gamma_1 &= 2\cosh\left(\frac{\lambda}{2}\right) \\
\gamma_2 &= 4+2\cosh\left(\lambda\right)
\end{aligned}
$$

$$
M\left(t\right) = \frac{1-e^{-\lambda}}{1-e^{t-\lambda}}
$$

$$
h\left[X\right] = \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} - \log\left(1-e^{-\lambda}\right)
$$

5 Poisson

The Poisson random variable counts the number of successes in n independent Bernoulli trials in the limit as $n \to \infty$ and $p \to 0$ where the probability of success in each trial is p and $np = \lambda \geq 0$ is a constant. It can be used to approximate the Binomial random variable or in it's own right to count the number of events that occur in the interval $[0, t]$ for a process satisfying certain "sparsity" constraints. The functions are

$$
\begin{aligned}
p\left(k;\lambda\right) &= e^{-\lambda}\frac{\lambda^k}{k!} \quad k \geq 0, \\
F\left(x;\lambda\right) &= \sum_{n=0}^{\lfloor x \rfloor} e^{-\lambda}\frac{\lambda^n}{n!} = \frac{1}{\Gamma\left(\lfloor x \rfloor + 1\right)}\int_{\lambda}^{\infty} t^{\lfloor x \rfloor} e^{-t}dt, \\
\mu &= \lambda \\
\mu_2 &= \lambda \\
\gamma_1 &= \frac{1}{\sqrt{\lambda}} \\
\gamma_2 &= \frac{1}{\lambda}.
\end{aligned}
$$

$$
M\left(t\right) = \exp\left[\lambda\left(e^t - 1\right)\right].
$$

6 Geometric

The geometric random variable with parameter $p \in (0, 1)$ can be defined as the number of trials required to obtain a success where the probability of success on each trial is p. Thus,

$$
\begin{aligned}
p\left(k;p\right) &= \left(1-p\right)^{k-1}p \quad k \geq 1 \\
F\left(x;p\right) &= 1-\left(1-p\right)^{\lfloor x \rfloor} \quad x \geq 1 \\
G\left(q;p\right) &= \left\lceil \frac{\log\left(1-q\right)}{\log\left(1-p\right)} \right\rceil \\
\mu &= \frac{1}{p} \\
\mu_2 &= \frac{1-p}{p^2} \\
\gamma_1 &= \frac{2-p}{\sqrt{1-p}} \\
\gamma_2 &= \frac{p^2-6p+6}{1-p}.
\end{aligned}
$$

$$M(t) = \frac{p}{e^{-t} - (1-p)}$$

7 Negative Binomial

The negative binomial random variable with parameters n and $p \in (0,1)$ can be defined as the number of *extra* independent trials (beyond n) required to accumulate a total of n successes where the probability of a success on each trial is p. Equivalently, this random variable is the number of failures encoutered while accumulating n successes during independent trials of an experiment that succeeds with probability p. Thus,

$$\begin{aligned}
p(k;n,p) &= \binom{k+n-1}{n-1} p^n (1-p)^k \quad k \geq 0 \\
F(x;n,p) &= \sum_{i=0}^{\lfloor x \rfloor} \binom{i+n-1}{i} p^n (1-p)^i \quad x \geq 0 \\
&= I_p(n, \lfloor x \rfloor + 1) \quad x \geq 0 \\
\mu &= n\frac{1-p}{p} \\
\mu_2 &= n\frac{1-p}{p^2} \\
\gamma_1 &= \frac{2-p}{\sqrt{n(1-p)}} \\
\gamma_2 &= \frac{p^2 + 6(1-p)}{n(1-p)}.
\end{aligned}$$

Recall that $I_p(a,b)$ is the incomplete beta integral.

8 Hypergeometric

The hypergeometric random variable with parameters (M,n,N) counts the number of "good" objects in a sample of size N chosen without replacement from a population of M objects where n is the number of "good" objects in the total population.

$$\begin{aligned}
p(k;N,n,M) &= \frac{\binom{n}{k}\binom{M-n}{N-k}}{\binom{M}{N}} \quad N-(M-n) \leq k \leq \min(n,N) \\
F(x;N,n,M) &= \sum_{k=0}^{\lfloor x \rfloor} \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}, \\
\mu &= \frac{nN}{M} \\
\mu_2 &= \frac{nN(M-n)(M-N)}{M^2(M-1)} \\
\gamma_1 &= \frac{(M-2n)(M-2N)}{M-2}\sqrt{\frac{M-1}{nN(M-m)(M-n)}} \\
\gamma_2 &= \frac{g(N,n,M)}{nN(M-n)(M-3)(M-2)(N-M)}
\end{aligned}$$

where (defining $m = M - n$)

$$\begin{aligned} g(N,n,M) &= m^3 - m^5 + 3m^2n - 6m^3n + m^4n + 3mn^2 \\ &\quad -12m^2n^2 + 8m^3n^2 + n^3 - 6mn^3 + 8m^2n^3 \\ &\quad +mn^4 - n^5 - 6m^3N + 6m^4N + 18m^2nN \\ &\quad -6m^3nN + 18mn^2N - 24m^2n^2N - 6n^3N \\ &\quad -6mn^3N + 6n^4N + 6m^2N^2 - 6m^3N^2 - 24mnN^2 \\ &\quad +12m^2nN^2 + 6n^2N^2 + 12mn^2N^2 - 6n^3N^2. \end{aligned}$$

9 Zipf (Zeta)

A random variable has the zeta distribution (also called the zipf distribution) with parameter $\alpha > 1$ if it's probability mass function is given by

$$p(k;\alpha) = \frac{1}{\zeta(\alpha)\,k^\alpha} \quad k \geq 1$$

where

$$\zeta(\alpha) = \sum_{n=1}^{\infty} \frac{1}{n^\alpha}$$

is the Riemann zeta function. Other functions of this distribution are

$$\begin{aligned} F(x;\alpha) &= \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\lfloor x \rfloor} \frac{1}{k^\alpha} \\ \mu &= \frac{\zeta_1}{\zeta_0} \quad \alpha > 2 \\ \mu_2 &= \frac{\zeta_2\zeta_0 - \zeta_1^2}{\zeta_0^2} \quad \alpha > 3 \\ \gamma_1 &= \frac{\zeta_3\zeta_0^2 - 3\zeta_0\zeta_1\zeta_2 + 2\zeta_1^3}{\left[\zeta_2\zeta_0 - \zeta_1^2\right]^{3/2}} \quad \alpha > 4 \\ \gamma_2 &= \frac{\zeta_4\zeta_0^3 - 4\zeta_3\zeta_1\zeta_0^2 + 12\zeta_2\zeta_1^2\zeta_0 - 6\zeta_1^4 - 3\zeta_2^2\zeta_0^2}{\left(\zeta_2\zeta_0 - \zeta_1^2\right)^2}. \end{aligned}$$

$$M(t) = \frac{\mathrm{Li}_\alpha\left(e^t\right)}{\zeta(\alpha)}$$

where $\zeta_i = \zeta(\alpha - i)$ and $\mathrm{Li}_n(z)$ is the n^{th} polylogarithm function of z defined as

$$\mathrm{Li}_n(z) \equiv \sum_{k=1}^{\infty} \frac{z^k}{k^n}$$

$$\mu_n' = M^{(n)}(t)\Big|_{t=0} = \frac{\mathrm{Li}_{\alpha-n}\left(e^t\right)}{\zeta(a)}\Big|_{t=0} = \frac{\zeta(\alpha-n)}{\zeta(\alpha)}$$

10 Logarithmic (Log-Series, Series)

The logarimthic distribution with parameter p has a probability mass function with terms proportional to the Taylor series expansion of $\log(1-p)$

$$\begin{aligned} p(k;p) &= -\frac{p^k}{k\log(1-p)} \quad k \geq 1 \\ F(x;p) &= -\frac{1}{\log(1-p)}\sum_{k=1}^{\lfloor x\rfloor}\frac{p^k}{k} = 1 + \frac{p^{1+\lfloor x\rfloor}\Phi(p,1,1+\lfloor x\rfloor)}{\log(1-p)} \end{aligned}$$

where

$$\Phi(z,s,a) = \sum_{k=0}^{\infty}\frac{z^k}{(a+k)^s}$$

is the Lerch Transcendent. Also define $r = \log(1-p)$

$$\begin{aligned} \mu &= -\frac{p}{(1-p)r} \\ \mu_2 &= -\frac{p[p+r]}{(1-p)^2 r^2} \\ \gamma_1 &= -\frac{2p^2+3pr+(1+p)r^2}{r(p+r)\sqrt{-p(p+r)}}r \\ \gamma_2 &= -\frac{6p^3+12p^2r+p(4p+7)r^2+\left(p^2+4p+1\right)r^3}{p(p+r)^2}. \end{aligned}$$

$$\begin{aligned} M(t) &= -\frac{1}{\log(1-p)}\sum_{k=1}^{\infty}\frac{e^{tk}p^k}{k} \\ &= \frac{\log(1-pe^t)}{\log(1-p)} \end{aligned}$$

Thus,

$$\mu_n' = M^{(n)}(t)\Big|_{t=0} = \frac{\mathrm{Li}_{1-n}(pe^t)}{\log(1-p)}\Bigg|_{t=0} = -\frac{\mathrm{Li}_{1-n}(p)}{\log(1-p)}.$$

11 Discrete Uniform (randint)

The discrete uniform distribution with parameters(a,b) constructs a random variable that has an equal probability of being any one of the integers in the half-open range $[a,b)$. If a is not given it is assumed to be zero and the only parameter is b. Therefore,

$$\begin{aligned} p(k;a,b) &= \frac{1}{b-a} \quad a \leq k < b \\ F(x;a,b) &= \frac{\lfloor x\rfloor - a}{b-a} \quad a \leq x \leq b \\ G(q;a,b) &= \lceil q(b-a)+a\rceil \\ \mu &= \frac{b+a-1}{2} \\ \mu_2 &= \frac{(b-a-1)(b-a+1)}{12} \\ \gamma_1 &= 0 \\ \gamma_2 &= -\frac{6}{5}\frac{(b-a)^2+1}{(b-a-1)(b-a+1)}. \end{aligned}$$

$$\begin{aligned} M(t) &= \frac{1}{b-a}\sum_{k=a}^{b-1} e^{tk} \\ &= \frac{e^{bt}-e^{at}}{(b-a)(e^{t}-1)} \end{aligned}$$

12 Discrete Laplacian

Defined over all integers for $a > 0$

$$\begin{aligned} p(k) &= \tanh\left(\frac{a}{2}\right) e^{-a|k|}, \\ F(x) &= \begin{cases} \frac{e^{a(\lfloor x\rfloor+1)}}{e^{a}+1} & \lfloor x\rfloor < 0, \\ 1-\frac{e^{-a\lfloor x\rfloor}}{e^{a}+1} & \lfloor x\rfloor \geq 0. \end{cases} \\ G(q) &= \begin{cases} \left\lceil \frac{1}{a}\log\left[q\left(e^{a}+1\right)\right]-1\right\rceil & q < \frac{1}{1+e^{-a}}, \\ \left\lceil -\frac{1}{a}\log\left[(1-q)\left(1+e^{a}\right)\right]\right\rceil & q \geq \frac{1}{1+e^{-a}}. \end{cases} \end{aligned}$$

$$\begin{aligned} M(t) &= \tanh\left(\frac{a}{2}\right)\sum_{k=-\infty}^{\infty} e^{tk}e^{-a|k|} \\ &= C\left(1+\sum_{k=1}^{\infty} e^{-(t+a)k}+\sum_{1}^{\infty} e^{(t-a)k}\right) \\ &= \tanh\left(\frac{a}{2}\right)\left(1+\frac{e^{-(t+a)}}{1-e^{-(t+a)}}+\frac{e^{t-a}}{1-e^{t-a}}\right) \\ &= \frac{\tanh\left(\frac{a}{2}\right)\sinh a}{\cosh a-\cosh t}. \end{aligned}$$

Thus,

$$\mu_n' = M^{(n)}(0) = \left[1+(-1)^n\right]\mathrm{Li}_{-n}\left(e^{-a}\right)$$

where $\mathrm{Li}_{-n}(z)$ is the polylogarithm function of order $-n$ evaluated at z.

$$h[X] = -\log\left(\tanh\left(\frac{a}{2}\right)\right)+\frac{a}{\sinh a}$$

13 Discrete Gaussian*

Defined for all μ and $\lambda > 0$ and k

$$p(k;\mu,\lambda) = \frac{1}{Z(\lambda)}\exp\left[-\lambda(k-\mu)^2\right]$$

where

$$Z(\lambda) = \sum_{k=-\infty}^{\infty}\exp\left[-\lambda k^2\right]$$

$$\begin{aligned} \mu &= \mu \\ \mu_2 &= -\frac{\partial}{\partial\lambda}\log Z(\lambda) \\ &= G(\lambda)e^{-\lambda} \end{aligned}$$

where $G(0)\to\infty$ and $G(\infty)\to 2$ with a minimum less than 2 near $\lambda = 1$

$$G(\lambda) = \frac{1}{Z(\lambda)}\sum_{k=-\infty}^{\infty} k^2\exp\left[-\lambda(k+1)(k-1)\right]$$