Families of Discrete Distributions

October 1, 2009

We shall, in general, denote the mass function of a parametric family of discrete distributions by $f_X(x|\theta)$ for the distribution depending on the parameter θ.

1 Discrete Uniform Distributions

X is said to have a **discrete uniform** $(1, N)$ distribution if the mass function

$$f_X(x|N) = \frac{1}{N}, \quad x = 1, 2, \ldots, N.$$

As we have seen,

$$EX = \frac{N+1}{2} \qquad \mathrm{Var}(X) = \frac{N^2-1}{12}.$$

The probability generating function

$$\rho_X(t) = \frac{1}{N}(z + z^2 + \cdots + z^N) = \frac{1}{N} \cdot \frac{z(1-z^N)}{1-z}.$$

Exercise 1. *Find the mean, variance and probability generating for a uniform (a, b) random variable.*

2 Bernoulli Distributions

X is said to have a *Bernoulli*(p) distribution if the mass function

$$f_X(x|p) = \begin{cases} 1-p & \text{if } \ x = 0, \\ p & \text{if } x = 1. \end{cases}$$

$$EX = p \qquad \mathrm{Var}(X) = p(1-p).$$

The probability generating function

$$\rho_X(t) = (1-p) + pz$$

3 Binomial Distributions

The binomial random variable is the number of successes in n Bernoulli trials. Its mass function is

$$f_X(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Previous computations have shown us that

$$EX = np, \qquad \mathrm{Var}(X) = np(1-p), \qquad M_X(t) = ((1-p)+pz)^n.$$

4 Hypergeometric Distributions

Consider an urn with B blue balls and G green balls. Remove K and let the random variable X denote the number of blue balls. Then the value of X is at most the maximum of B and K. If $K > G$, then we might choose all of the green balls. If $X + x$, then the number of green balls $K - x \leq G$ and thus, $x \geq K - G$

The mass function for X is

$$f_X(x|B,G,K) = \frac{\binom{B}{x}\binom{G}{K-x}}{\binom{B+G}{K}}, \quad x = \max\{0, K-G\}, \ldots, \min\{B, K\}.$$

We can rewrite this as

$$f_X(x|B,G,K) = \frac{K!}{x!(K-x)!}\frac{(B)_x(G)_{K-x}}{(B+G)_K} = \binom{K}{x}\frac{(B)_x(G)_{K-x}}{(B+G)_K}.$$

This is an example of sampling **without replacement**. If we were to choose the balls one-by-one returning the balls to the urn after each choice, then we would be sampling **with replacement**. This returns us to the case of K Bernoulli trials with success parameter $p = G/(B+G)$. In the case the mass function for Y, the number of green balls is

$$f_Y(y|B,G,K) = \binom{K}{y}\frac{B^x G^{K-y}}{(B+G)^K}.$$

Let Y_i be a Bernoulli random variable indicating whether or not the color of the i-th is blue. Thus, its mean

$$EY_i = \frac{B}{B+G}.$$

The random variable $Y = Y_1 + Y_2 + \cdots + Y_K$ and thus its mean

$$EY = EY_1 + EY_2 + \cdots + EY_K = K\frac{B}{B+G}.$$

We will later be able to compute the variance

$$\mathrm{Var}(Y) = K\frac{B}{B+G} \cdot \frac{G(B+G-K)}{(B+G)(B+G-1)}.$$

If we write $N = B + G$ and p as above, then

$$\mathrm{Var}(Y) = Kp(1-p)\frac{N-K}{(N-1)}$$

and thus the variance is reduced by a factor of $(N-K)/(N-1)$ from the case of a binomial random variable.

5 Poisson Distributions

The Poisson distribution is an approximation of the binomial distribution in the can that n is large and p is small, but the product $\lambda = np$ is moderate. In this case

$$\begin{aligned}
P\{X=0\} &= \binom{n}{0}p^0(1-p)^n &= \left(1-\frac{\lambda}{n}\right)^n &\approx e^{-\lambda} \\
P\{X=1\} &= \binom{n}{1}p^1(1-p)^{n-1} &= n\frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)^{n-1} &\approx \lambda e^{-\lambda} \\
P\{X=2\} &= \binom{n}{2}p^2(1-p)^{n-2} &= \frac{(n)_2}{2}\left(\frac{\lambda}{n}\right)^2\left(1-\frac{\lambda}{n}\right)^{n-2} &\approx \frac{\lambda^2}{2}e^{-\lambda} \\
\vdots \quad & \quad \vdots & \vdots \quad & \\
P\{X=x\} &= \binom{n}{x}p^x(1-p)^{n-x} &= \frac{(n)_x}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} &\approx \frac{\lambda^x}{x!}e^{-\lambda}
\end{aligned}$$

because

$$\frac{(n)_x}{n^x} \approx 1$$

.

A random variable X has a **Poisson distribution** if its mass function

$$f_X(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}.$$

The Taylor series expansion for $\exp \lambda$ show that $\sum_x f_X(x|\lambda) = 1$. The generating function

$$\rho_X(z) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}e^{-\lambda}z^x = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda z)^x}{x!} = e^{-\lambda}e^{\lambda z} = \exp \lambda(z-1).$$

To check the approximation of a binomial distribution by a Poisson, note that

x	$Pois(1)$	$Bin(10, 0.1)$	$Bin(100, 0.01)$
0	0.3679	0.3487	0.3660
1	0.3679	0.3874	0.3697
2	0.1839	0.1937	0.1849
3	0.0613	0.0574	0.0610
4	0.0153	0.0112	0.0149
5	0.0031	0.0015	0.0029
6	0.0005	0.0001	0.0005

Exercise 2. *Show that $EX = \lambda$ and $Var(X) = \lambda$.*

Exercise 3. *Let ρ_n be the generating function for a binomial random variable based on n trials with success probability λ/n. Show that*

$$\lim_{n\to\infty} \rho_n(z) = \rho(z),$$

the generating function for a Poisson random variable with parameter λ.

6 Geometric Distributions

The geometric random variable is the time of the first success in a sequence of Bernoulli trials.

$$f(x|p) = p^{x-1}(1-p), \qquad x = 1, 2, \ldots.$$

For this random variable, we have

$$EX = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}, \quad \rho_X(z) = \frac{pz}{1-(1-p)z}.$$

7 Negative Binomial Distributions

The number of failures before of the r-th success is called a **negative binomial** random variable. to determine its mass function, note that

$$\begin{aligned} P\{X = x\} &= P\{r-1 \text{ successes in } x+r-1 \text{ trials and success on the } x+r\text{-th trial}\} \\ &= P\{r-1 \text{ successes in } x+r-1 \text{ trials}\} \cdot P\{\text{success on the } x+r\text{-th trial}\} \\ &= \binom{x+r-1}{r-1} p^{r-1}(1-p)^x \cdot p = \binom{x+r-1}{x} p^r (1-p)^x. \end{aligned}$$

The generating function

$$\rho_X(z) = \sum_{x=0}^{\infty} \binom{x+r-1}{x} p^r (1-p)^x z^x = p^r \sum_{x=0}^{\infty} \frac{(x+r-1)_x}{x!} \alpha^r = (1-\alpha)^{-r}$$

where $\alpha = (1-p)z$, i.e., $\rho_X(z) = (1-(1-p)z)^{-r}$.

Exercise 4. *Check that the Taylor's series expansion of $g(\alpha) = (1-\alpha)^{-r}$ is the infinite sum given above. This gives the power series expansion of a negative power of the binomial. For this reason, X is called a* **negative binomial distribution**.

Exercise 5. *Show that*

$$EX = \frac{r}{p}, \qquad \mathrm{Var}(X) = r\frac{1-p}{p^2}$$