Chapter 2 - Discrete Distributions

Reminder: A discrete random variable has space R consisting of a finite or countable infinite set of values.

Defn: A **probability mass function** (pmf) of a random variable of discrete type satisfies the following properties:

- f(x)>0, for all $x \in R$. (The probability of an outcome that may occur is positive.)
- $\sum_{x \in R} f(x) = 1$. (An outcome from the sample space always occurs.)
- For A a subset of R, P($X \in A$) = $\sum_{x \in A} f(x)$. (Probs. are additive.)
- f(x)=0, for all $x \notin R$. (The probability of something that cannot occur is zero.)

Note, this makes the domain of f the entire real numbers and not just the subset R. Since f is zero except for values in R, we often call R the *support* of f.

Defn: The distribution function F(x) corresponding to a random variable X and its pmf f(x) is given by:

$$F(x) = \sum_{t \le x} f(t) .$$

Practically, F gives a running total of the probabilities as the random variable increases.

Result: If X is a discrete random variable, we get the following properties for a distribution function:

- $0 \le F(x) \le 1$.
- F(x) is a nondecreasing function.
- F(y) = 1 for any value y larger than the largest value in R.
- F(y) = 0 for any value y smaller than the smallest value in R.
- F is a step function with the jump at x=a equal to P(X=a).

Graphing the Distribution: For a discrete probability distribution, we simply plot the points (x,f(x)), for all $x \in R$. To get a better picture of the distribution, we use *bar graphs* and *histograms*. A bar graph is simply a set of lines connecting (x,0) and (x,f(x)). If X takes on only integer values, then a histogram is a collection of rectangles centered at each $x \in R$ of height f(x) and width one. A probability histogram illustrates the probability associated with each discrete x $x \in R$ by associating the corresponding amount of area around x with f(x).

Uniform Distribution: A discrete random variable X has a *uniform distribution* on the integers R = {1,2,...,n} if each outcome is equally likely. Then, one obtains:

X({outcome #x}) = x

f(x) = 1/n is constant

$\mu = (n+1)/2$

$\sigma^2 = (n^2-1)/12$.

Notice, on some problems, the space is R = {0, 1, ..., n} and this changes the results above slightly.

Hypergeometric Distribution: Suppose data falls into 2 classes, n_1 in the first and n_2 in the second, $n=n_1+n_2$. Select without replacement a collection of r items from a random mixture of the objects and let X be the random variable giving the number of items in the sample from the first class. Notice, $0 \le x \le n_1$, $0 \le x \le n_2$ and $0 \le x \le r$. Then,

$$N(X=x) = \binom{n_1}{x}\binom{n_2}{r-x}$$

and since the sample is randomly chosen, using the uniform distribution gives

$$f(x) = P(X=x) = \binom{n_1}{x}\binom{n_2}{r-x} \Big/ \binom{n}{r} .$$

which is called the pmf of the *hypergeometric distribution*.

Result: f(x) as above satisfies the properties of a pdf.x

Pf: By the Binomial Theorem, $(1+y)^n = \sum_{x=0}^{n}\binom{n}{x}y^x$.

Therefore, $\sum_{x=0}^{n}\binom{n}{x}y^x = (1+y)^n = (1+y)^{n_1}(1+y)^{n_2} = \sum_{x=0}^{n_1}\binom{n_1}{x}y^x \times \sum_{x=0}^{n_2}\binom{n_2}{x}y^x$.

and multiplying yields $\sum_{x=0}^{n}\binom{n}{x}y^x = \sum_{x=0}^{n}\sum_{t=0}^{r}\binom{n_1}{t}\binom{n_2}{r-t}y^x$,

where for notational purposes we set $\binom{a}{b}=0$ for b>a.

By equating like powers of y, we get for the rth power term *(Equation 2.1-JT)*

$$\binom{n}{x}y^x = \sum_{t=0}^{r}\binom{n_1}{t}\binom{n_2}{r-t}y^x$$

By dividing, we get 1 = $\sum_{t=0}^{n} f(x)$, as desired.

HOMEWORK: page 66

Defn: Assume f(x) is a pmf of the discrete random variable X, define the *expected value* or *mathematical expectation* of the function u(x) to be

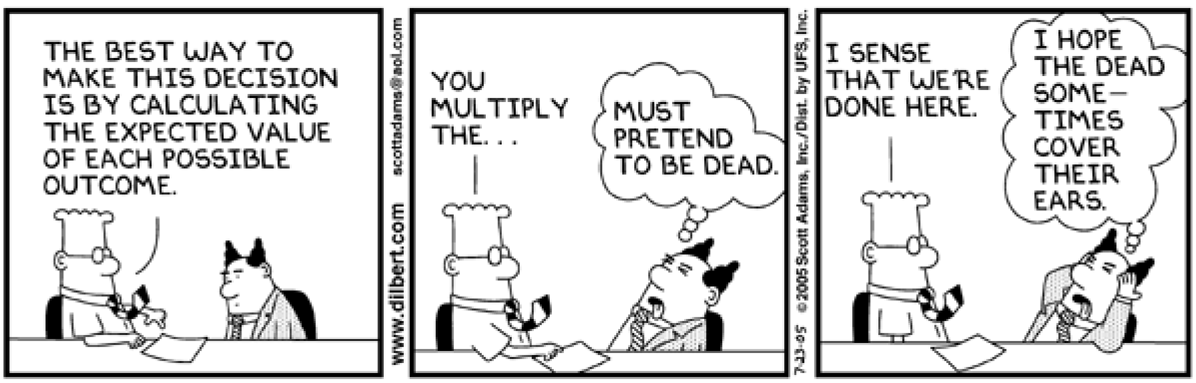
$$E = E[u(X)] = \sum_{x \in R} u(x) f(x),$$

if the sum converges. We often call f(x) the weight of u(x).

Note: If X is finite, the sum always converges. Also, if each x has equal weight, f(x) = 1/n for all x.
So, E = Σ u(x)/n = average of u(x) values. If X is countable infinite, the sum might also be infinite.

Result: Expected value is a linear operator

- E[c]=c, for c a constant. (You expect a constant to be that constant.)
- ii) E[c u(X)] = c E[u(X)], for c a constant. (You can factor out constants.)
- iii) E[c_1u_1(X) + c_2u_2(X)] = c_1E[u_1(X)] + c_2E[u_2(X)], for c_1, c_2 constants & u_1, u_2 functions.



HOMEWORK: page 130, all problems

Defn: Suppose X is a discrete random variable with range R = $\{x_1, x_2, ..., x_n\}$ and pmf f(x).

- The *mean* of X is given by

$$\mu = E(X) = x_1 f(x_1) + ... + x_n f(x_n).$$

Notice, if X is "equally likely", f(x) = 1/n and in this case yields $\mu = E(X) = \Sigma f(x) / n$.

The mean is often called the *centroid* in the sense that if the x_k were locations of objects of weight $f(x_k)$, then the centroid would be the point where this system of n masses would balance.

- The *variance* is defined by

$$E[(X-\mu)^2],$$

if this value exists and is denoted by σ^2 or Var(X).

- The positive square root of the variance is called the *standard deviation* and is denoted by σ.
- The *coefficient of skewness* for a distribution is given by

$$\mu_3 = E[(X-\mu)^3] / \sigma^3$$

if this value exists. This yields a measure which indicates the symmetry of the distribution.

- The *kurtosis* for a distribution is given by

$$\mu_4 = E[(X-\mu)^4] / \sigma^4$$

if this value exists. This yields a measure which indicates the flatness or peakedness of the distribution.

Result: $\sigma^2 = E[X^2] - \mu^2 = E[X(X-1)] + \mu - \mu^2$

(This gives two possibly easier ways of computing σ^2 and therefore σ.)

Defn: If $E[X^r]$ exists, it is called the *rth moment* of the distribution about the origin and $E[(X-b)^r]$ is called the rth moment of the distribution about b.

$$E[(X)_r] = E[X(X-1)...(X-r+1)]$$

is the *rth factorial moment*. Note, $E[(X)_1] = E[X(X-1)] = E[X^2] - \mu^2$.

Defn: If we have n observations of the random variable X (ie. a sample), then we get empirical results. In particular the mean of these observations is the mean of an empirical distribution and is denoted $\bar{x}$. Generally, the empirical mean $\bar{x}$ is very close to μ = E[X]. In much of statistics, we try to quantify how close they are. The empirical mean is most valuable if μ is unknown (as it often is) in that we can approximate μ by $\bar{x}$. In a similar vein, we define the empirical variance denoted v. Often statisticians redefine the empirical variance by

$$s^2 = \frac{\sum_{k=1}^{m} (x_k - \bar{x})^2}{n-1} = v\, n / (n-1).$$

One reason for this is to make the variance slightly larger to compensate for “bias” where extreme values are often left out when sampling. This makes s^2 a better estimate of σ^2 than v is. However, for large sample sizes, we see that s^2 and v are practically identical.

Lemma: $\binom{n}{r} = \frac{n}{r}\binom{n-1}{r-1}$

Result: If X is the random variable of a hypergeometric distribution, then

$$\mu = r\, n_1/n, \text{ and}$$

$$\sigma^2 = n\, (n_1/n)\, (n_2/n)\, [(n-r)/(n-1)].$$

Pf: $\mu = E(X) = \sum_{x=0}^{r} x\, f(x) = \sum_{x=0}^{r} x \binom{n_1}{x}\binom{n_2}{r-x} \Big/ \binom{n}{r}$, and by using the Lemma and canceling,

$= n_1 \sum_{x=1}^{r} \binom{n_1 - 1}{x-1}\binom{n_2}{r-x} \Big/ \binom{n}{r}$; setting s = r - 1 and j = x - 1 yields

$= n_1 \sum_{j=0}^{s} \binom{n_1 - 1}{j}\binom{n_2}{s-j} \Big/ \binom{n}{s+1}$; considering n_1-1 replacing n_1 and s replacing r in Equation 2.1-JT yields

$= n_1 \binom{n_1 + n_2 - 1}{s} \Big/ \binom{n}{s+1}$; using the lemma above again and canceling terms leaves us with,

$= r\dfrac{n_1}{n}$.

The proof of the variance formula is similar and uses the result $\sigma^2 = E(X(X-1) + \mu - \mu^2$. The proof of skewness and kurtosis are messy and we won't bother with them for this distribution!

HOMEWORK: page 84

Bernoulli Distribution: Let X be a random variable mapping success to 1 and failure to 0. That is,

X(success)=1 and X(failure)=0.

Suppose P(success) = p is known. Then, the Bernoulli Distribution is given by

$$f(x) = p^x(1-p)^{1-x},$$

$$\mu = p,$$

$$\sigma^2 = p\,(1-p).$$

Binomial Distribution: Let X be a random variable measuring the number of successes in a succession of a fixed number n of independent Bernoulli trials. Notice, the space of X is R = { 0, 1, 2,..., n }. To have x successes implies n-x failures. So, there are $\binom{n}{x}$ ways of getting x successes and each, by independence, has probability $p^x(1-p)^{n-x}$. So, the Binomial distribution is given by

$$f(x) = \binom{n}{x} p^x(1-p)^{n-x},$$

$$\mu = np,$$

$$\sigma^2 = n\,p\,(1-p).$$

Notation: b(n,p)

Result: f(x) is a pmf

Pf: By the binomial theorem, $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$.

Apply with a = p and b = 1-p to get $1 = (p+1-p)^n = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x}$.

Result: The binomial distribution mean is np.

Pf: $\mu = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} = np \sum_{t=0}^{m} \binom{m}{t} p^t (1-p)^{m-t} = np$.

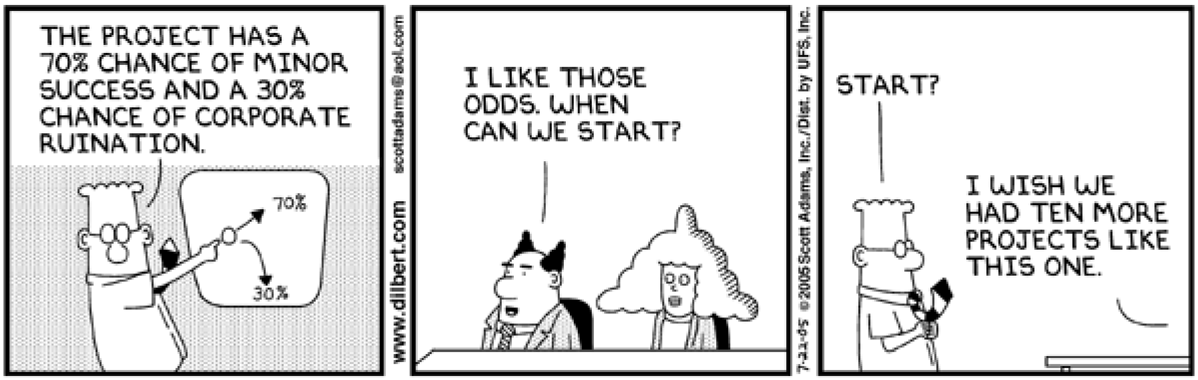
Result: The binomial distribution variance is np(1-p).

Pf:
$$\begin{aligned}
\sigma^2 &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x (1-p)^{n-x} + np - n^2p^2 \\
&= n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} (1-p)^{n-x} + np - n^2p^2 \\
&= n(n-1)p^2 \sum_{t=0}^{m} \binom{m}{t} p^t (1-p)^{m-t} + np - n^2p^2 \\
&= n(n-1)p^2 + np - n^2p^2 \\
&= -np^2 + np \\
&= np(-p+1)
\end{aligned}$$

Result: The binomial distribution coefficient of skewness is $\frac{1-2p}{\sqrt{np(1-p)}}$. Therefore, the binomial distribution is symmetrical when p = ½.

Result: The binomial distribution kurtosis is $\frac{1-6p(1-p)}{np(1-p)}+3$. Therefore, the binomial distribution is bell-shaped provided p(1-p) ≈ 1/6.

Remark: The distribution function F(x) has no nice form and can be computed using the tables in the book or by using the Statistical Features of your graphing calculator: Usage: F(x) = binomcdf(n,p,x).



If the Pointy-Haired Boss has his way, compute the Probability of Corporate Ruination…

HOMEWORK: page 97

Geometric Distribution: Let X be a random variable measuring the number of independent Bernoulli trials it takes to get one success. Since there must have been x-1 failures to begin with, then

$$f(x)=(1-p)^{x-1}p,$$
$$\mu=1/p$$
$$\sigma^2=(1-p)/p^2.$$

Result: The geometric distribution coefficient of skewness is $\frac{2-p}{\sqrt{1-p}}$. Therefore, the geometric distribution is always skewed to the right.

Result: The geometric distribution kurtosis is $\frac{p^2-6p+6}{1-p}+3=\frac{p^2}{1-p}+9$ Therefore, the geometric distribution is not at all bell-shaped.

Remark: To verify that f(x) is a pmf, one uses the Geometric series from Calculus III.

Remark: The distribution function F(x) has a nice form when using the Geometric distribution. Indeed:

Indeed, $P(X > x) = f(x+1) + f(x+2) + \ldots = \sum_{t=x+1}^{\infty}(1-p)^{t-1}p = p(1-p)^x\sum_{u=0}^{\infty}(1-p)^u = p(1-p)^x\frac{1}{1-(1-p)} = (1-p)^x$

Therefore, using complements, $F(x) = 1 - (1-p)^x$.

Also, one may use geometriccmf(p,x) on a graphing calculator.

Negative Binomial: Let X be a random variable measuring the number of independent Bernoulli trials it takes to get a fixed number r of successes. For r = 1, this is the Geometric distribution. For r > 1, in the x-1 trials before the last one, there must have been r-1 successes and x-r failures.

This can happen $\binom{x-1}{r-1}$ ways each of which has probablity $(1-p)^{x-r}p^{r-1}$. Therefore,

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$$
$$\mu = r/p$$
$$\sigma^2 = r(1-p)/p^2.$$

Notice, we can get a formula for F(x) when r=2 using the derivative of a geometric series but in general this if difficult.

Result: The negative binomial distribution coefficient of skewness is $\frac{2-p}{\sqrt{r(1-p)}}$. Therefore, the negative binomial distribution is always skewed to the right.

Result: The negative binomial distribution kurtosis is $\frac{p^2-6p+6}{r(1-p)}+3=\frac{p^2}{1-p}+\frac{6+3r}{r}$ Therefore, the negative binomial distribution is not generally bell-shaped but gets close as r increases provided $p \approx 0$.

Poisson Distribution: An approximate Poisson process satisfies the following:

- the number of changes in disjoint intervals is independent
- the probability of exactly one change in a sufficiently short interval of length h is proportional to h
- the probability of two or more changes in the above interval is essentially zero.

Let X measure the <u>number of changes</u> in a Poisson process on a given interval. Notice, the interval from which the changes may occur is not likely discrete but the number of changes in the interval will be discrete. To determine the mass function, divide the interval into n>>x equal subintervals of width t/n, where t = width of the interval. Then,

$$p = P(\text{exactly one change in a subinterval}) = \lambda(t/n)$$

and since the probability of two or more is essentially zero, on a given interval the there are two options—success (a change) or failure (no change). If we consider each sub-interval as a Bernoulli trial, we have a sequence of n Bernoulli trials with fixed n and p = λ(t/n), which is a Binomial problem.
Therefore, for the Poisson variable X,

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

To get an exact value, fix x and let n approach infinity to get for the Poisson distribution

$$f(x) = (t\lambda)^x e^{-(t\lambda)}/x!$$

$$\mu = t\lambda$$

$$\sigma^2 = t\lambda.$$

Result: The poisson distribution coefficient of skewness is $\frac{1}{\sqrt{\lambda}}$. Therefore, the poisson distribution is always skewed to the right but becomes more symmetrical as λ increases

Result: The poisson distribution kurtosis is $\frac{1}{\lambda}+3$ Therefore, the poisson distribution is relatively bell-shaped for larger values of λ.

Note: Since the binomial approaches the poisson as n gets large, then one may also use the Poisson distribution to approximate the binomial.

Computational: F(x) = poissoncdf(μ,x)

HOMEWORK: page 121