4.E.1. The Binomial Distribution

Consider a random experiment of two outcomes 0 & 1 with corresponding probabilities

$$p_0 = q \qquad \& \qquad p_1 = p \tag{4.38a}$$

such that

$$p + q = 1 \tag{4.38b}$$

Consider now the case where in N trials, 0 occurs n_0 times & 1, n_1 times, with $n_0 + n_1 = N$.

During each run, the 0's & 1's can occur in any one of 2^N possible sequences. The probability of a particular sequence to occur is $q^{n_0} p^{n_1}$. The number of sequences having the same n_1 is given by the number of ways to put n_1 black balls and $N - n_1$ white balls into N boxes, namely, $C^N_{n_1} = C^N_{N-n_1}$. By the rule of addition, the probability of finding the stated outcome is therefore

$$P_N(n_1) = C^N_{n_1}\, q^{n_0}\, p^{n_1} = \frac{N\,!}{n_1\,!\,(N - n_1)\,!}\, q^{N-n_1}\, p^{n_1} \tag{4.38}$$

Result for the case of $N = 3$ is tabulated as follows

sequences	n_1	n_0
000	0	3
001	1	2
010	1	2
100	1	2
011	2	1
101	2	1
110	2	1
111	3	0

n_1	$C^3_{n_1}$
0	1
1	3
2	3
3	1

(4.38) is known as the binomial distribution because $C^N_{n_1}$ is the coefficient of p^{n_1} in the binomial expansion of $(p + q)^N$.

Using the binomial theorem, we get the normalization condition

$$\sum_{n_1=0}^{N} P_N(n_1) = \sum_{n_1=0}^{N} C^N_{n_1}\, p^{n_1}\, q^{N-n_1} = (p+q)^N = 1 \qquad [\text{ (4.38b) used. }] \tag{4.39}$$

For the jth trial, the probability assignment (4.38a) corresponds to the probability density

$$P_{X_j}(x_j) = q\,\delta(x_j) + p\,\delta(x_j - 1) \tag{4.39a}$$

where X_j is the stochastic variable for the jth trial and x_j its realization. The corresponding characteristic function is [see (4.15)]

$$f_{X_j}(k_j) = \int_{-\infty}^{\infty} d\,x_j\, e^{i\,k_j\,x_j}\,[\,q\,\delta(x_j) + p\,\delta(x_j - 1)\,]$$

$$= q + p\, e^{i\,k_j} \tag{4.39b}$$

Let each sequence of N trials be represented by the stochastic variable

$$Y^{(N)} = \sum_{j=1}^{N} X_j \qquad \to \qquad y^{(N)} = \sum_{j=1}^{N} x_j = n_1 \tag{4.40a}$$

Since $\{X_j\}$ are statistically independent,

$$P_{Y^{(N)}}(y) = P_{X_1}(x_1) \ldots P_{X_N}(x_N) \tag{4.40b}$$

so that (4.34) gives

$$P_{Y^{(N)}}(y) = \int d\,x_1 \ldots \int d\,x_N\, \delta(y - x_1 - \ldots - x_N)\, P_{X_1}(x_1) \ldots P_{X_N}(x_N) \qquad (4.40)$$

where, since there is no ambiguity here, we have replaced $y^{(N)}$ with y for the sake of clarity.

Putting in (4.39a), we have

$$P_{Y^{(N)}}(y) = \int d\,x_2 \ldots \int d\,x_N \left[\, \delta\,(y - x_2 - \ldots - x_N)\; q + \delta\,(y - 1 - x_2 - \ldots - x_N)\, p \,\right] P_{X_2}(x_2) \ldots P_{X_N}(x_N)$$

$$= \int d\,x_3 \ldots \int d\,x_N \left[\, \delta\,(y - x_3 - \ldots - x_N)\; q^2 + 2\,\delta\,(y - 1 - x_3 - \ldots - x_N)\, p\, q \right.$$

$$\left. + \,\delta\,(y - 2 - x_3 - \ldots - x_N)\, p^2 \,\right] P_{X_3}(x_3) \ldots P_{X_N}(x_N)$$

$$\vdots$$

$$= \int d\,x_{m+1} \ldots \int d\,x_N \sum_{n_1=0}^{m} C^m_{n_1}\, q^{N-m}\, p^m\, \delta\,(y - n_1 - x_{m+1} - \ldots - x_N)\, P_{X_{m+1}}(x_{m+1}) \ldots P_{X_N}(x_N)$$

$$\vdots$$

$$= \sum_{n_1=0}^{N} C^N_{n_1}\, \delta(y - n_1)\; q^{N-n_1}\, p^{n_1}$$

$$= \sum_{n_1=0}^{N} P_N(n_1)\, \delta(y - n_1) \qquad [\ (4.38)\text{ used.}\ ] \qquad (4.43)$$

Alternatively, with $P_{Y^{(N)}}(y)$ taking the form (4.40b), $f_{Y^{(N)}}(k)$ also assumes a simple product form. Putting (4.40 & b) into (4.35), we get

$$f_{Y^{(N)}}(k) = \int d\,x_1 \ldots \int d\,x_N\, e^{i\,k\,(x_1 + \ldots + x_N)}\, P_{X_1}(x_1) \ldots P_{X_N}(x_N)$$

$$= f_{X_1}(k) \ldots f_{X_N}(x_N) \qquad [\ (4.15)\text{ used.}\ ]$$

$$= \left(q + p\, e^{i\,k}\right)^N \qquad [\ (4.39\text{b})\text{ used.}\ ] \qquad (4.41)$$

$$= \sum_{n_1=0}^{N} C^N_{n_1}\, q^{N-n_1}\, p^{n_1}\, e^{i\,k\,n_1} \qquad [\ \text{Binomial expansion.}\ ] \qquad (4.42)$$

which can also be obtained by putting (4.43) into (4.15) so that

$$f_{Y^{(N)}}(k) = \int d\,y\, e^{i\,k\,y} \sum_{n_1=0}^{N} P_N(n_1)\, \delta(y - n_1)$$

$$= \sum_{n_1=0}^{N} P_N(n_1)\, e^{i\,k\,n_1}$$

$$= \sum_{n_1=0}^{N} C^N_{n_1}\, q^{N-n_1}\, p^{n_1}\, e^{i\,k\,n_1} \qquad [\ (4.38)\text{ used.}\ ]$$

Taking the inverse Fourier transform of (4.15), we have

$$P_{Y^{(N)}}(y) = \int_{-\infty}^{\infty} \frac{d\,k}{2\,\pi}\, e^{-i\,k\,y}\, f_{Y^{(N)}}(k)$$

$$= \sum_{n_1=0}^{N} C^N_{n_1}\, q^{N-n_1}\, p^{n_1} \int_{-\infty}^{\infty} \frac{d\,k}{2\,\pi}\, e^{-i\,k\,(y - n_1)} \qquad [\ (4.42)\text{ used.}\ ]$$

$$= \sum_{n_1=0}^{N} C^N_{n_1}\, q^{N-n_1}\, p^{n_1}\, \delta(y - n_1)$$

which is just (4.43).

The moments for $Y^{(N)}$ are given by

$$\langle y^m \rangle = \langle n_1^m \rangle \qquad \text{[(4.40a) used.]}$$
$$= \sum_{n_1=0}^{N} n_1^m P_N(n_1)$$
$$= \sum_{n_1=0}^{N} n_1^m C_{n_1}^N q^{N-n_1} p^{n_1}$$
$$= \left(p \frac{\partial}{\partial p}\right)^m \sum_{n_1=0}^{N} C_{n_1}^N q^{N-n_1} p^{n_1}$$
$$= \left(p \frac{\partial}{\partial p}\right)^m \sum_{n_1=0}^{N} P_N(n_1)$$
$$= \left(p \frac{\partial}{\partial p}\right)^m (p+q)^N \qquad (4.44a)$$

The same result can also be obtained using (4.17) & (4.41)

$$\langle y^m \rangle = (-i)^m \left.\frac{\partial^m f_{Y^{(N)}}(k)}{\partial k^m}\right|_{k=0}$$
$$= (-i)^m \left.\frac{\partial^m \left(q + p\, e^{ik}\right)^N}{\partial k^m}\right|_{k=0} \qquad \text{[(4.41) used.]} \qquad (4.44b)$$

Thus,

$$\langle n_1 \rangle = \left(p \frac{\partial}{\partial p}\right)(p+q)^N = N p (p+q)^{N-1} = N p \qquad \text{[(4.44a) used.]} \qquad (4.44)$$
$$= (-i) \left.\frac{\partial \left(q + p\, e^{ik}\right)^N}{\partial k}\right|_{k=0} = N p\, e^{ik} \left(q + p\, e^{ik}\right)^{N-1} \Big|_{k=0} = N p \qquad \text{[(4.44b) used.]} \qquad (4.45)$$
$$\left\langle n_1^2 \right\rangle = \left(p \frac{\partial}{\partial p}\right)^2 (p+q)^N = \left(p \frac{\partial}{\partial p}\right)\left[N p (p+q)^{N-1} \right] \qquad \text{[(4.44a) used.]}$$
$$= N p \left[(p+q)^{N-1} + (N-1)\; p (p+q)^{N-2} \right]$$
$$= N p \left[1 + (N-1)\; p \right]$$
$$= N p (N p + q)$$
$$= (N p)^2 + N p q \qquad (4.46)$$
$$= N p (-i) \left.\frac{\partial\, e^{ik} \left(q + p\, e^{ik}\right)^{N-1}}{\partial k}\right|_{k=0} \qquad \text{[(4.45) used.]}$$
$$= N p \left[1 + (N-1) p \frac{e^{ik}}{q + p\, e^{ik}} \right] e^{ik} \left(q + p\, e^{ik}\right)^{N-1} \Big|_{k=0}$$
$$= N p \left[1 + (N-1) p \right]$$

$$\rightarrow \quad \mathrm{Var}\left(Y^{(N)}\right) = \left\langle n_1^2 \right\rangle - \langle n_1 \rangle^2$$
$$= N p q \qquad \text{[(4.45-6) used.]} \qquad (4.47a)$$

$$\therefore \quad \sigma_{Y^{(N)}} = \sqrt{N p q} \qquad (4.47)$$

The fractional deviation (or standard deviation in units of the mean) is

$$\frac{\sigma_{Y(N)}}{\langle y \rangle} = \frac{\sqrt{N p q}}{N p} = \sqrt{\frac{q}{N p}} \tag{4.48}$$

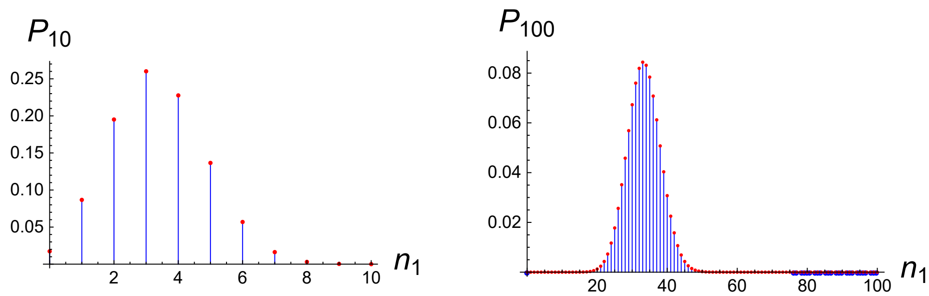

Fig.4.2. Binomial distributions for $p = \frac{1}{3}$, $N = 10$ and $N = 100$.

Code

```
In[]:= P[n1_, n_ , p_] := n! / (n1! (n - n1)!) p^n1 (1 - p)^(n-n1)
```

```
In[]:= (* Fig.4.2. *)
n = 10; p = 1/3;
ListPlot[ Table[{n1, P[n1, n, p]}, {n1, 0, n}],
      AxesLabel → {"n1", "P10" }, PlotStyle → Red,
      Filling → Bottom, FillingStyle → {Thick, Blue}
]
```

```
In[]:= (* Fig.4.2. *)
n = 100; p = 1/3;
ListPlot[ Table[{n1, P[n1, n, p]}, {n1, 0, n}],
      PlotRange → All,
      AxesLabel → {"n1", "P100" }, PlotStyle → Red,
      Filling → Bottom, FillingStyle → {Thick, Blue}
]
```