distribution may be suitable for random variables with bell shape distributions within finite boundaries. Examples are cost distributions, repair time distributions, and distributions of fluctuating reservoir levels (Yevjevich, 1972, p. 149). The pdf's given by Equations (2.7-30) through (2.7-34) integrated over the range $0 \le t \le 1$ produce 1, as required.

Some of the *moments* of the beta distribution can be rather easily obtained by integration. For example, for Equation (2.7-34), the first moment is obtained by

$$\mu = \int_0^1 t f(t) dt = 30 \int_0^1 t^3 (1-t)^2 dt = 30[\frac{t^4}{4} - 2\frac{t^5}{5} + \frac{t^6}{6}]|_0^1 = \frac{30}{4} - \frac{60}{5} + \frac{30}{6} = \frac{1}{2} .$$

The mean is at the center of the range for this symmetric pdf.

Formulas for moments of the beta distribution can be found in handbooks (Moan, 1982, pp. 4-40 to 4-41). For example, the *k*-th moment of the beta function is

$$\nu_k = \frac{(\alpha+\beta+1)!(\alpha+k)!}{(\alpha+\beta+k+1)!(\alpha)!} \tag{2.7-35}$$

For the mean, or first moment, $k = 1$, and with $\alpha = 2$, and $\beta = 2$,

$$\nu_1 = \mu = \frac{5!\,3!}{6!\,2!} = \frac{1}{2} .$$

This is the same result as obtained by the integration of Equation (2.7-34).

Examples: (1) Show that Equation (2.7-30) is a pdf. Solution: The integral of Equation (2.7-30) over the range of the random variable $0 \le t \le 1$ is

$$\int_0^1 f(t) dt = \frac{1}{\pi} \int_0^1 \frac{1}{\sqrt{t}} \frac{dt}{\sqrt{1-t}}$$

We use the substitution

$u = \sqrt{1-t}$, which also is $t = 1 - u^2$. Introducing $-2\,du = \dfrac{dt}{\sqrt{1-t}}$

and the substitution of $\sqrt{t}$ into the integral gives

$$-\frac{2}{\pi}\int \frac{du}{\sqrt{1-u^2}} = -\frac{2}{\pi}\arcsin(u) = -\frac{2}{\pi}(\arcsin\sqrt{1-t})\Big|_0^1 = -\frac{2}{\pi}(0-\frac{\pi}{2}) = 1 \quad .$$

This proves that Equation (2.7-30) is a pdf.

(2) For Equation (2.7-30), calculate the mean by the moment method and also by Equation (2.7-35). Solution: The moment method proceeds similar to the integration in Example (1). The first moment is

$$\mu = \int_0^1 t\, f(t)dt = \frac{1}{\pi}\int_0^1 \frac{t}{\sqrt{t}}\frac{dt}{\sqrt{1-t}} \quad .$$

With the same substitutions as in Example (1) the integral is transformed and solved for the transformed integration limits: for $t = 0$, $u = 1$; and for $t = 1$, $u = 0$. The result is

$$\mu = -\frac{2}{\pi}\int \sqrt{1-u^2}\,du = -\frac{2}{\pi}\left[\frac{u}{2}\sqrt{1-u^2} + \frac{1}{2}\arcsin(u)\right]_{u=1}^{u=0} = \frac{1}{2} \quad .$$

The mean, computed by Equation (2.7-35), for $k = 1$, $\alpha = -1/2$, and $\beta = -1/2$, requires the evaluation of several factorials: $0! = 1$, $(1/2)! = (1/2)(\pi/2)$, $1! = 1$; and $(-1/2)! = \pi/2$ (see the factorial function in Abramowitz and Stegun, 1970, p. 255). Substituting these results into Equation (2.7-35) one obtains $\mu = 1/2$, the same as by the moment method. Equation (2.7-30) is a symmetric pdf and the mean is located in the center of the range.

2.7.7 Poisson Distribution

The Poisson distribution is used to model arrival rates and service rates in queuing models. There is a connection between the *exponential* and the *Poisson* distributions. Probabilistic (or stochastic) queuing models use exponential distributions for interarrival time and service time modeling or, equivalently, the Poisson distribution for arrival rate and service rate modeling (Gross and Harris, 1974, p. 23). The number of arrivals during a period extending from time zero to t is a random variable. The probability of n arrivals during a period t, with $n \geq 0$ being positive integers, is derived by setting up so-called stochastic *differential-difference equations*. This calculus sheds light on probabilistic mathematics, but the derivation is somewhat involved and cannot be repeated here. Suffice it to say that the results of the derivation are the *stochastic differential equations* of arrival probabilities (Gross and Harris, 1974, p. 25). For zero arrivals,

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t), \tag{2.7-36a}$$

where the index relates to the number of arrivals in time t. For arrivals $n \geq 1$, the stochastic differential equation is of the form

$$\frac{dp_n(t)}{dt} = -\lambda p_n(t) + \lambda\, p_{n-1}(t)\,. \tag{2.7-36b}$$

Equation (2.7-36b) is a first order linear differential equation in which the dependent variable and its derivative are of the first degree. A solution method consists of finding an *integrating factor* (Nielsen, 1962, p. 48). First Equation (2.7-36b) is rewritten in the form

$$\frac{dp_n(t)}{dt} + \lambda p_n(t) = \lambda\, p_{n-1}(t)\,. \tag{2.7-36c}$$

Suppose an integrating factor, $g(t)$, has been found. Then, multiplying the equation with it gives

$$g(t)[\frac{dp_n(t)}{dt} + \lambda p_n(t)] = g(t)\lambda\, p_{n-1}(t)\ . \tag{2.7-36d}$$

The integrating factor is of such a nature that it complements the left-hand side of the equation into an exact differential. The integral of an exact differential is readily obtained whereupon only the right-hand side needs to be integrated. The probability $p_{n-1}(t)$ is a constant that is known from a previous calculation so that the right-hand side can also be easily integrated. A hint of an integrating factor is obtained by the solution of the homogeneous equation, Equation (2.7-36a),

$$\frac{dp_n(t)}{dt} + \lambda p_n(t) = 0\,.$$

The solution is obtained by separation of variables and integration,

$$p_n(t) = ce^{-\lambda t}$$

where c is an integration constant. Moving the e-function to the left side gives

$$e^{\lambda t} p_n(t) = c\,.$$

Taking the total differential of this equation gives

$$d[e^{\lambda t} p_n(t)] = e^{\lambda t}[\frac{dp_n(t)}{dt} + \lambda p_n(t)] = 0$$

This shows that $e^{\lambda t}$ is an integrating factor that complements the left side of Equation (2.7-36d). With $g(t) = e^{\lambda t}$, the left-hand side of Equation (2.7-36d) can be written as an exact differential $d[e^{\lambda t} p_n(t)]$ and the integral of the equation becomes

$$\int_0^t d[e^{\lambda t} p_n(t)] = \int_0^t e^{\lambda t} \lambda\, p_{n-1}(t) dt \quad . \tag{2.7-36e}$$

An equation explicit in the dependent variable $p_n(t)$ is obtained by carrying out the integration of the left-hand side. The product of functions on the left-hand side exists only for $t > 0$. Freeing $p_n(t)$ of its factor gives

$$p_n(t) = e^{-\lambda t} \int_0^t e^{\lambda t} \lambda\, p_{n-1}(t) dt \tag{2.7-36f}$$

where the factor in front of the integral is the result of the integration of the left-hand side. Equation (2.7-36f) is valid for $n \geq 1$. It is a recursive formula which can be used to find the probabilities for $n > 0$ once $p_0(t)$ is found.

For $n = 0$, which means zero arrivals during period t, Equation (2.7-36a) applies. By separation of variables one obtains

$$p_0(t) = e^{-\lambda t} \tag{2.7-37a}$$

where λ is the mean arrival rate of items per time unit, and $\lambda\, t$ is the mean number of arrivals in period t.

For $n = 1$, one obtains from Equation (2.7-36f) by integrating from 0 to t,

$$p_1(t) = e^{-\lambda t} \int_0^t e^{\lambda t} \lambda\, p_0(t) dt = e^{-\lambda t} \int_0^t e^{\lambda t} \lambda\, e^{-\lambda t} dt = \lambda\, t\, e^{-\lambda t} \quad . \tag{2.7-37b}$$

For $n = 2$: $$p_2(t) = \frac{(\lambda t)^2}{2} e^{-\lambda t} \quad . \qquad (2.7\text{-}37c)$$

For $n = 3$: $$p_3(t) = \frac{(\lambda t)^3}{3!} e^{-\lambda t} \quad . \qquad (2.7\text{-}37d)$$

The probability of $n = k$ arrivals in period t is

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad . \qquad (2.7\text{-}38)$$

Equation (2.7-38) is the pdf of the Poisson arrival process. It is a discrete probability distribution, where $p_k(t)$ is the discrete probability of $k = 0, 1, 2, 3, \ldots, k$ arrivals in period t. The Poisson distribution approaches the binomial distribution for large n and small p, with $\lambda t = n p$. The random variable does not have to be time. If λ is the number of accidents per kilometer of highway, then the average number of accidents over 10 km is $a = 10 \lambda$.

The parameter $a = \lambda t$ in Equation (2.7-38) is the only parameter of the Poisson distribution. It is also the mean as well as the standard deviation of the distribution: $\mu = \sigma = a$. The mean can be interpreted as expected number of jobs arriving at a repair shop during a specified period t, whereas the pdf gives the frequencies of random realizations that may actually occur causing job queues. Two examples of Equation (2.7-38) for $a = 5$ and $a = 10$ are shown in Figure 2-24.

The non-exceedance probability of the Poisson distribution is

$$F(k) = e^{-\lambda t} \sum_{i=0}^{k} \frac{(\lambda t)^i}{i!} \quad . \qquad (2.7\text{-}39)$$

where $F(k)$ is the discrete cdf of the Poisson distribution whose ordinates only exist for integer numbers.

A stochastic process with practical applications is the *compound Poisson process*. A practical example is the sum of costs that accrue by the random occurrence of repair jobs during some time period, such as a month (Parzen, 1965, p. 128),

$$X(t) = \sum_{n=1}^{N(t)} C_n$$

where $X(t)$ is the stochastic variable, i.e., the total of $N(t)$ random costs, C_n, that occur during a month t by a Poisson arrival process of repair jobs.

Examples: (1) Calculate the probability of $n = 4$ arrivals in period t with $p_3(t)$ given by Equation (2.7-37d). Solution: Using Equation (2.7-36f), one obtains

$$p_4(t) = e^{-\lambda t} \int_0^t e^{\lambda t} \lambda \, p_3(t) \, dt = e^{-\lambda t} \int_0^t e^{\lambda t} \lambda \frac{(\lambda t)^3}{3!} e^{-\lambda t} dt = \frac{(\lambda t)^4}{4!} e^{-\lambda t} .$$

(2) What is the probability that up to and including six arrivals occur in period $t =$ 10 h, given the average arrival rate is $\lambda = 0.6 \text{ h}^{-1}$. Solution: The average number of arrivals is $\lambda t = 0.6 \cdot 10 = 6$. The non-exceedance probability of six arrivals is according to Equation (2.7-39)

$$F(6) = p_0(10) + p_1(10) + p_2(10) + p_3(10) + p_4(10) + p_5(10) + p_6(10) =$$

$$= (6^0/0!)\, e^{-6} + (6^1/1!)\, e^{-6} + (6^2/2!)\, e^{-6} + (6^3/3!)\, e^{-6}$$

$$+ (6^4/4!)\, e^{-6} + (6^5/5!)\, e^{-6} + (6^6/6!)\, e^{-6}$$

$$= 0.00248\,(1 + 6 + 18 + 36 + 54 + 64.8 + 64.8) = 0.6066.$$

(3) The binomial cdf, Equation (2.7-8), is adapted to Example (2) to find the probability of 6 successes in 10 trials, or the probability of 6 arrivals (successes) in 10 hours ($n = 10$ trials). For $n = 10$, $k = 6$, $p = 6/10 = 0.6$, Equation (2.7-8) becomes

$$F(6) = \sum_{i=0}^{6} [B(10,i) \cdot 0.6^i \cdot 0.4^{10-i}]$$

$$= B(10,0) \cdot 0.6^0 \cdot 0.4^{10} + B(10,1) \cdot 0.6^1 \cdot 0.4^9 + B(10,2) \cdot 0.6^2 \cdot 0.4^8$$

$$+ B(10,3) \cdot 0.6^3 \cdot 0.4^7 + B(10,4) \cdot 0.6^4 \cdot 0.4^6 + B(10,5) \cdot 0.6^5 \cdot 0.4^5$$

$$+ B(10,6) \cdot 0.6^6 \cdot 0.4^4$$

$$= 1 \cdot 1 \cdot 0.000\,105 + 10 \cdot 0.6 \cdot 0.000\,262 + 45 \cdot 0.36 \cdot 0.000\,655$$

$$+ 120 \cdot 0.216 \cdot 0.001\,64 + 210 \cdot 0.129\,6 \cdot 0.004\,096$$

$$+ 252 \cdot 0.077\,76 \cdot 0.01024 + 210 \cdot 0.046\,66 \cdot 0.025\,6$$

$$= 0.618 \approx 0.62$$

This result was also obtained by Example (4) of Section 2.7.2.

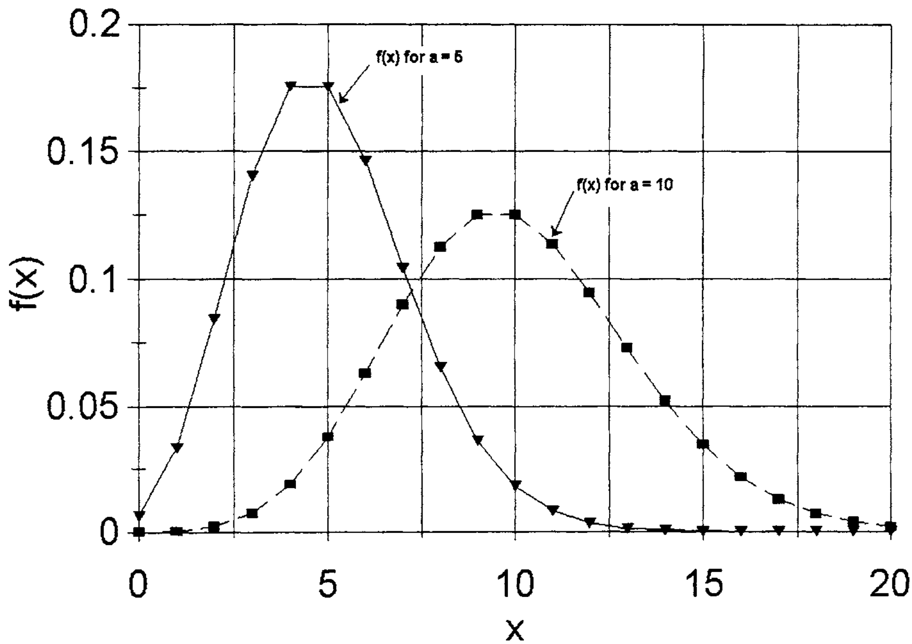

Figure 2-24: Poisson pdf's for random arrivals k in the form of $f(x) = a^x e^{-a}/x!$ where $a = \lambda t$, and $x = k$ for average arrivals $\mu = \lambda t = 5$ and 10. The arrival probabilities are highest around the mean. The random variable $x = k$ is a discrete variable which exists only for integers $x \geq 0$. The $f(x)$ are discrete probabilities, which should be shown as bars over the x's for which they exist. The line connections between the $f(x)$ represent only visual aids.

(4) Calculate the probability of up to two defective items arriving in a production stream over a 50 h period if the average defect rate is $\lambda = 0.01\ h^{-1}$. Compare the result with the binomial distribution. Solution: The probability of up to two defective items is given by $F(2)$. For $\lambda t = 0.01 \cdot 50 = 0.5$,

$$F(2) = p_0(50) + p_1(50) + p_2(50)$$

$$= (0.5^0/0!)\, e^{-0.5} + (0.5^1/1!)\, e^{-0.5} + (0.5^2/2!)\, e^{-0.5}$$

$$= 0.6065 + 0.3033 + 0.0758 = 0.9856.$$

This result is very close to the first part of Example (7) of Section 2.7.2, which produced $F(2) = 0.9862$. It demonstrates that the Poisson distribution approaches the binomial distribution for large n and small p. The Poisson probabilities are easier to calculate than the binomial probabilities.

2.7.8 Weibull Distribution

The Weibull distribution was first proposed for "a statistical representation of fatigue failures in solids" (Moan, 1982, p. 4-47), but it also is used in hydrology because it is applicable for positive only random variables (Stedinger et al., 1993, p. 18.13). It is a three-parameter distribution and therefore is quite flexible in the shapes it can simulate. Various formulas are given in the numerous references. Here, the pdf is given as referred to by Parzen (1965, p. 169):

$$f(x) = \frac{k}{v-c}\left(\frac{x-c}{v-c}\right)^{k-1} e^{-\left(\frac{x-c}{v-c}\right)^k} \qquad (2.7\text{-}40)$$

where $f(x)$ exists only for $x \geq c$, and is zero for $x < c$; c is the location parameter, k is the shape parameter, and v is the scale parameter. Equation (2.7-40) looks somewhat complex, but it is actually quite simple and can be easily integrated to obtain the non-exceedance probability. For this purpose, we introduce the substitution

$$u = \left(\frac{x-c}{v-c}\right)^k .$$

The differential of u is

$$du = k\left(\frac{x-c}{v-c}\right)^{k-1} \frac{dx}{v-c} .$$

Substituting u and du into the integral of the cdf, one obtains

$$F(x) = \int_c^x f(x)dx = -\int_c^x e^{-u} d(-u) = -e^{-u}\Big|_c^x .$$

Evaluating the boundaries gives

$$F(x) = 1 - e^{-(\frac{x-c}{v-c})^k} . \tag{2.7-41}$$

The simple relation between $f(x)$ and $F(x)$ allows easy use of the Weibull distribution in the calculation of the hazard function (Section 3.6).

A graphical determination of parameters is based on taking the double logarithm of a form derived from Equation (2.7-41):

$$\frac{1}{1 - F(x)} = e^{(\frac{x-c}{v-c})^k} .$$

This leads to

$$\ln(\ln \frac{1}{1 - F(x)}) = k \ln(x - c) - k \ln(v - c) . \tag{2.7-42}$$

Equation (2.7-42) is a straight line in a double-log diagram:

$$Y = k\,X + C \tag{2.7-43}$$

where

$$Y = \ln(\ln \frac{1}{1 - F(x)}) , \tag{2.7-44}$$

$$X = \ln(x - c) , \tag{2.7-45}$$

and

$$C = -k \ln(v - c) . \tag{2.7-46}$$

where Y and X are the coordinates of the double log diagram, C is the empirical constant or intercept, and k is the slope of the line (Moan, 1982, p. 4-49).

Special Cases of the Weibull distribution: The general form of the three-parameter Weibull distribution, Equation (2.7-40), can be simplified for special cases:

(1) If the distribution starts at $x = 0$, then $c = 0$, and Equation (2.7-40) takes the form

$$f(x) = \frac{k}{\nu}\left(\frac{x}{\nu}\right)^{k-1} e^{-\left(\frac{x}{\nu}\right)^k}, \tag{2.7-47}$$

and

$$F(x) = 1 - e^{-\left(\frac{x}{\nu}\right)^k} \tag{2.7-48}$$

where x, ν, and $k > 0$. This form is frequently used. Its mean and variance are

$$\mu = \nu\,\Gamma\left(1 + \frac{1}{k}\right) \tag{2.7-48a}$$

and

$$\sigma^2 = \nu^2 \left\{\Gamma\left(1 + \frac{2}{k}\right) - \left[\Gamma\left(1 + \frac{1}{k}\right)\right]^2\right\} \tag{2.7-48b}$$

where $\Gamma(.)$ is the gamma function (Stedinger et al., 1993, p. 18.13; Benjamin and Cornell, 1970, p. 284). For $k = 2$, $\Gamma(3/2) = 0.5\ \pi^{0.5} = 0.88623$, and for integers, $\Gamma(n+1) = n!$ For $n = 1$, $\Gamma(2) = 1$. Hence, for $k = 2$, $\mu = 0.8862\ \nu$, and $\sigma^2 = 0.2146\ \nu^2$. Both μ and σ^2 are proportional to ν and ν^2, respectively, as is shown by Equations (2.7-48a) and (2.7-48b), with the gamma functions providing the proportionality coefficients. The parameters k and ν can be found from a data fit using Equations (2.7-44) to (2.7-46). According to Equation (2.7-43), k is the slope of the line fitting the data. More on gamma functions is found in Abramowitz and Stegun (1970, p. 255).

(2) For $k = 1$, and $c = 0$, $\lambda = 1/\nu$, Equation (2.7-40) becomes

$$f(x) = \lambda e^{-\lambda x} \tag{2.7-49a}$$

and

$$F(x) = 1 - e^{-\lambda x}. \tag{2.7-49b}$$