Chapter 6
Some Continuous Probability Distributions

吳明龍 副教授
成大資訊系/醫資所
minglong.wu@csie.ncku.edu.tw
辦公室:資訊系館 12 樓

6.5 Normal Approximation to the Binomial

- Poisson distribution can be used to approximate binomial probabilities when *n* is quite large and *p* is very close to 0 or 1.
- Normal distribution not only provide a very accurate approximation to binomial distribution when *n* is large and *p* is not extremely close to 0 or 1, but also provides a fairly good approximation even when *n* is small and *p* is reasonably close to ½.
- Theorem 6.3: If *X* is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$, then the limiting form of the distribution of

$$Z = \frac{X - np}{\sqrt{npq}}$$

as $n \to \infty$, is the standard normal distribution $n(z;0,1)$

Normal Approximation to the Binomial

- Normal approximation to the binomial distribution

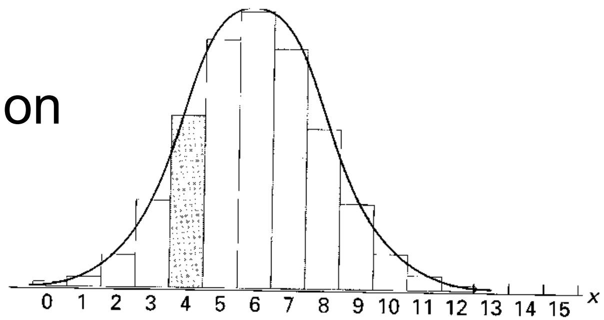

Figure 6.23 Normal approximation of $b(4; 15, 0.4)$ and $\sum_{x=7}^{9} b(x; 15, 0.4)$.

$$(1)\ P(X = 4) = b(4;15,0.4) = 0.1268$$

$$\mu = np = 15 \cdot 0.4 = 6, \sigma^2 = 15 \cdot 0.4 \cdot 0.6 = 3.6, \sigma = 1.897$$

$$z_1 = \tfrac{3.5-6}{1.897} = \text{-}1.32 \text{ and } z_2 = \tfrac{4.5-6}{1.897} = \text{-}0.79$$

$$P(X = 4) \approx P(3.5 < X < 4.5) = P(-1.32 < Z < -0.79) = P(Z < -0.79) - P(Z < -01.32)$$

$$= 0.2148 - 0.0934 = 0.1214$$

$$(2)\ P(7 \le X \le 9) = \sum_{x=7}^{9} b(x;15,0.4) = \sum_{x=0}^{9} b(x;15,0.4) - \sum_{x=0}^{6} b(x;15,0.4)$$

$$= 0.9662 - 0.6098 = 0.3564$$

$$z_1 = \tfrac{6.5-6}{1.897} = 0.26 \text{ and } z_2 = \tfrac{9.5-6}{1.897} = 1.85$$

$$P(7 \le X \le 9) \approx P(0.26 < Z < 1.85) = P(Z < 1.85) - P(Z < 0.26)$$

$$= 0.9678 - 0.6026 = 0.3652$$

Normal Approximation to the Binomial

- The degree of accuracy, which depends on how well the curve fits the histogram, will increase as n increases.
- If both np and nq are greater than or equal to 5, the normal approximation will be good.

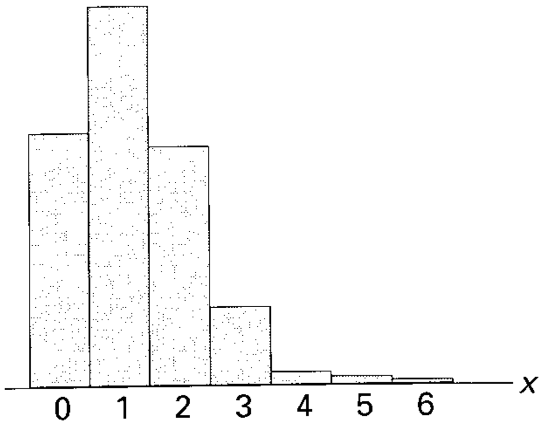

Figure 6.24 Histogram for $b(x; 6, 0.2)$.

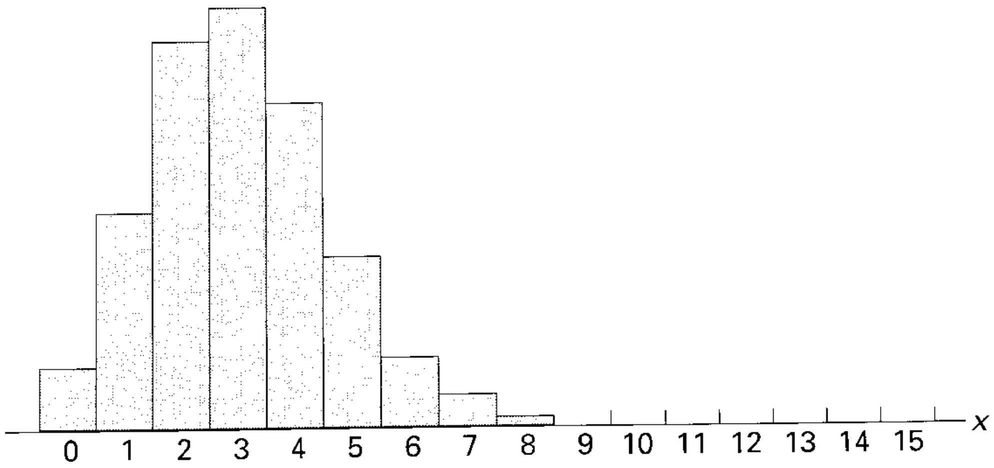

Figure 6.25 Histogram for $b(x; 15, 0.2)$.

Normal Approximation to the Binomial

- Example 6.15: The probability that a patient recovers from a rare blood disease is 0.4. If 100 people are known to have contracted this disease, what is the probability that less than 30 survive?
  - **Solution**

$$\mu = np = 100 \cdot 0.4 = 40, \sigma = \sqrt{100 \cdot 0.4 \cdot 0.6} = 4.899$$

$$z_1 = \tfrac{29.5-40}{4.899} = -2.14$$

$$P(X < 30) \approx P(Z < -2.14) = 0.0162$$

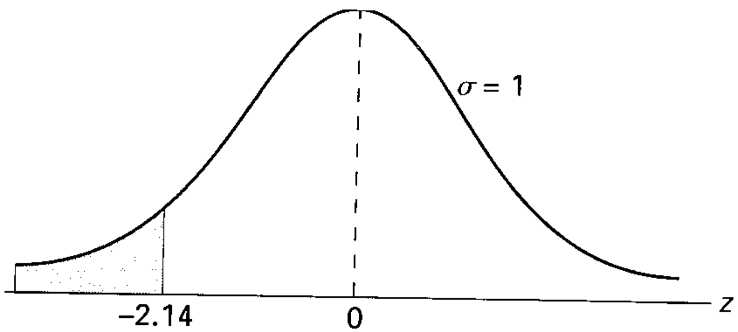

Figure 6.26 Area for Example 6.15.

Normal Approximation to the Binomial

- Example 6.16: A multiple-choice quiz has 200 questions each with 4 possible answers of which only 1 is correct answer. What is the probability that sheer (全然的) guess-work yields from 25 to 30 correct answers for 80 of the 200 problems about which the student has no knowledge?
  - **Solution**

$$\mu = np = 80 \cdot \tfrac{1}{4} = 20, \sigma = \sqrt{80 \cdot \tfrac{1}{4} \cdot \tfrac{3}{4}} = 3.873$$

$$z_1 = \tfrac{24.5-20}{3.873} = 1.16,\ z_2 = \tfrac{30.5-20}{3.873} = 2.71$$

$$P(25 \le X \le 30) = \textstyle\sum_{x=25}^{30} b(x; 80, \tfrac{1}{4})$$

$$\approx P(1.16 < Z < 2.71)$$

$$= P(Z < 2.71) - P(Z < 1.16)$$

$$= 0.9966 - 0.8770 = 0.1196$$

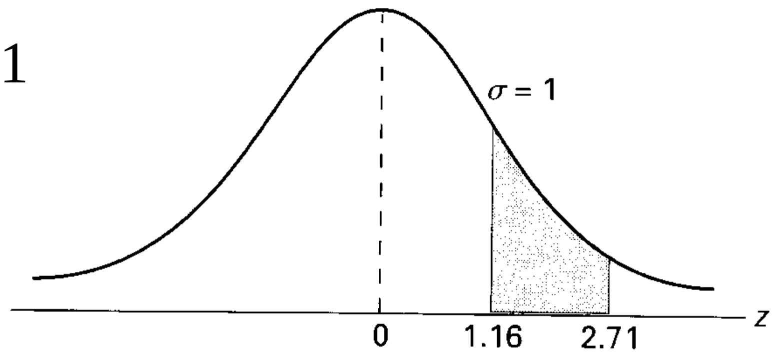

Figure 6.27 Area for Example 6.16.

6.6 Gamma and Exponential Distributions

- Exponential is a special case of the gamma distribution.
- Play an important role in queuing theory and reliability problems.
  - Time between arrivals at service facilities, time to failure of component parts and electrical systems.
- Definition 6.2: The gamma function is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \text{ for } \alpha > 0$$

$$\int u dv = uv - \int v du$$

Integrating by parts with $u = x^{\alpha-1}$ and $dv = e^{-x}dx$

$$\Gamma(\alpha) = -e^{-x} x^{\alpha-1} |_0^\infty + \int_0^\infty e^{-x} (\alpha-1) x^{\alpha-2} dx = (\alpha-1)\int_0^\infty x^{\alpha-2} e^{-x} dx, \text{ for } \alpha > 1$$

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$$

$$= (\alpha-1)(\alpha-2)\Gamma(\alpha-2) = (\alpha-1)(\alpha-2)(\alpha-3)\Gamma(\alpha-3) = \cdots$$

$$\Gamma(n) = (n-1)(n-2)\cdots\Gamma(1)$$

$$\because \Gamma(1) = \int_0^\infty e^{-x} dx = 1 \therefore \Gamma(n) = (n-1)!$$

Gamma and Exponential Distributions

- Gamma Distribution: The continuous random variable *X* has a gamma distribution, with parameters α and β, if its density function is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & \text{elsewhere,} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$.

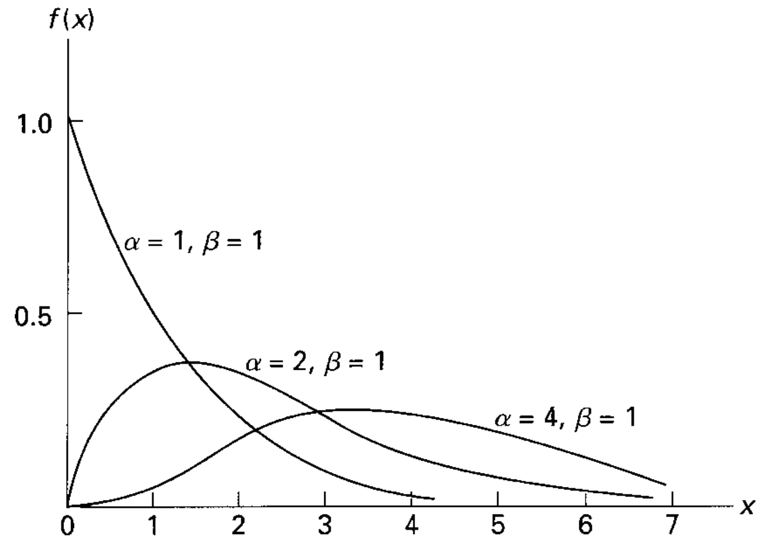

Figure 6.28 Gamma distributions.

Gamma and Exponential Distributions

- Exponential Distribution (α=1, special gamma distribution): The continuous random variable X has an exponential distribution, with parameters β, if its density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0, & \text{elsewhere,} \end{cases}$$

where $\beta > 0$.

- The mean and variance of the gamma distribution are

$$\mu = \alpha\beta \text{ and } \sigma^2 = \alpha\beta^2.$$

  – Proof is in Appendix A.26(A.28).

- The mean and variance of the exponential distribution are

$$\mu = \beta \text{ and } \sigma^2 = \beta^2.$$

Gamma and Exponential Distributions

A.26(A.28)

PROOF To find the mean of the gamma distribution, we write

$$\mu = E(X) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^\alpha e^{-x/\beta} dx.$$

Now, let $y = x/\beta$, to give

$$\mu = \frac{\beta}{\Gamma(\alpha)} \int_0^\infty y^\alpha e^{-y} dy = \frac{\beta \Gamma(\alpha + 1)}{\Gamma(\alpha)} = \alpha\beta.$$

To find the variance of the gamma distribution, we proceed as above to obtain

$$E(X^2) = \frac{\beta^2 \Gamma(\alpha + 2)}{\Gamma(\alpha)} = (\alpha + 1)\alpha\beta^2,$$

and then

$$\sigma^2 = E(X^2) - \mu^2 = (\alpha + 1)\alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2.$$

Gamma and Exponential Distributions

$$f(x) = \frac{1}{\beta} e^{-x/\beta}$$

- Relationship to the Poisson Process
  - The most important applications of the exponential distribution are situations where the Poisson process applies.
  - An industrial engineer may be interested in modeling the time *T* between arrivals at a congested interaction during rush hour in a large city. An arrival represents the Poisson event.
  - Using Poisson distribution, the probability of no events occurring in the span up to time *t*

$$p(0, \lambda t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

  - Let *X* be the time to the first Poisson event.
    - The probability that the length of time until the first event will exceed *x* is the same as the probability that no Poisson events will occur in *x*.

$$P(X \geq x) = e^{-\lambda x} \Rightarrow P(0 \leq X \leq x) = 1 - e^{-\lambda x}$$

  - Differentiate the cumulative distribution function to obtain the exponential distribution $f(x) = \lambda e^{-\lambda x}$ with $\lambda = \frac{1}{\beta}$

Applications of Gamma and Exponential Distributions

- The mean of the exponential distribution is the parameter β, the reciprocal (倒數) of the parameter λ in the Poisson distribution.
- Poisson distribution has no memory, implying that occurrences in successive time periods are independent.
- The important parameter β is the mean time between events. In reliability theory, equipment failure often conforms to this Poisson process, β is called mean time between failures.
- Many equipment breakdowns do follow the Poisson process, and thus the exponential distribution does apply. Other applications include survival times in bio-medical experiments and computer response time.

Applications of Gamma and Exponential Distributions

- Example 6.17: Suppose that a system contains a certain type of component whose time in years to failure is given by *T*. The random variable *T* is modeled nicely by the exponential distribution with mean time to failure $\beta = 5$. If 5 of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years.
  - Solution

$$P(T > 8) = \tfrac{1}{5}\int_8^\infty e^{-t/5}\,dt = e^{-8/5} \approx 0.2$$

Let X represent the number of components functioning after 8 years.

$$P(X \geq 2) = \sum_{x=2}^{5} b(x;5,0.2) = 1 - \sum_{x=0}^{1} b(x;5,0.2) = 1 - 0.7373 = 0.2627$$

Applications of Gamma and Exponential Distributions

- Example 6.18: Suppose that telephone calls arriving at a switchboard follow a Poisson process with an average of 5 calls coming per minute. What is the probability that up to a minute will occur until 2 calls have come in to the switchboard?
  - Solution

The Poisson process applies with time until 2 Poisson events following a gamma distribution with $\beta = 1/5$ and $\alpha = 2$.

Let X represent the time in minutes that occurs before 2 calls come.

$$P(X \le x) = \int_0^x \frac{1}{\beta^2} x e^{-x/\beta} dx$$

$$P(X \le 1) = 25\int_0^1 x e^{-5x} dx = 1 - e^{-5(1)}(1+5) = 0.96$$

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

Applications of Gamma and Exponential Distributions

- Example 6.19: In a biomedical study with rats a dose-response investigation is used to determine the effect of the dose of a toxicant (毒物) on their survival time. The toxicant is one that is frequently discharged into the atmosphere from jet fuel. For a certain dose of the toxicant the study determines that the survival time, in weeks, has a gamma distribution with α = 5 and β =10. what is the probability that a rat survives no longer than 60 weeks?
  - Solution

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

Let X be the survival time

$$P(X \le x) = \int_0^x \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx$$

$$P(X \le 60) = \frac{1}{\beta^5} \int_0^{60} \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(5)} dx$$

Using incomplete gamma function $F(x;\alpha) = \int_0^x \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy$

Let $y = x/\beta, x = \beta y$

$$\Rightarrow P(X \le 60) = \int_0^6 \frac{y^4 e^{-y}}{\Gamma(5)} dy = P(X \le 60) = F(6;5) = 0.715 \text{ (see Appendix A.23(A.24))}$$

Applications of Gamma and Exponential Distributions

TABLE A.24 The Incomplete Gamma Function: $F(x;\alpha) = \int_0^x \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy$

x \ α	1	2	3	4	5	6	7	8	9	10
1	0.632	0.264	0.080	0.019	0.004	0.001	0.000	0.000	0.000	0.000
2	0.865	0.594	0.323	0.143	0.053	0.017	0.005	0.001	0.000	0.000
3	0.950	0.801	0.577	0.353	0.185	0.084	0.034	0.012	0.004	0.001
4	0.982	0.908	0.762	0.567	0.371	0.215	0.111	0.051	0.021	0.008
5	0.993	0.960	0.875	0.735	0.560	0.384	0.238	0.133	0.068	0.032
6	0.998	0.983	0.938	0.849	0.715	0.554	0.394	0.256	0.153	0.084
7	0.999	0.993	0.970	0.918	0.827	0.699	0.550	0.401	0.271	0.170
8	1.000	0.997	0.986	0.958	0.900	0.809	0.687	0.547	0.407	0.283
9		0.999	0.994	0.979	0.945	0.884	0.793	0.676	0.544	0.413
10		1.000	0.997	0.990	0.971	0.933	0.870	0.780	0.667	0.542
11			0.999	0.995	0.985	0.962	0.921	0.857	0.768	0.659
12			1.000	0.998	0.992	0.980	0.954	0.911	0.845	0.758
13				0.999	0.996	0.989	0.974	0.946	0.900	0.834
14				1.000	0.998	0.994	0.986	0.968	0.938	0.891
15					0.999	0.997	0.992	0.982	0.963	0.930

Chi-Squared Distribution

- Chi-Squared Distribution ($\alpha = v/2$ and $\beta = 2$, special gamma distribution): The continuous random variable X has a chi-squared distribution, with v degrees of freedom, if its density function is given by

$$f(x) = \begin{cases} \dfrac{1}{2^{v/2}\Gamma(v/2)} x^{v/2-1} e^{-x/2}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

where v is a positive integer.

- The chi-squared distribution is an important component of statistical hypothesis testing and estimation.
- The mean and variance of the chi-squared distribution are

$$\mu = v \text{ and } \sigma^2 = 2v.$$

Lognormal Distribution

- The lognormal distribution applies in cases where a natural log transformation results in a normal distribution.
- Lognormal Distribution: The continuous random variable *X* has a lognormal distribution if the random variable *Y* = ln(*X*) has a normal distribution with mean μ and standard deviation σ. The resulting density function of X is

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}[(\ln(x)-\mu)/\sigma]^2}, & x \geq 0 \\ 0, & \mathrm{x} < 0 \end{cases}$$

- The mean and variance of the lognormal distribution are

$$E(X) = e^{\mu+\sigma^2/2} \text{ and } Var(\mathrm{X}) = e^{2\mu+\sigma^2} \cdot (e^{\sigma^2} - 1).$$

Lognormal Distribution

- Example 6.22: Concentration (濃度) of pollutants produced by chemical plants historically are known to exhibit behavior that resembles a lognormal distribution. This is important when one considers issues regarding compliance to government regulations. Suppose it is assumed that the concentration of a certain pollutant, in parts per million, has a lognormal distribution with parameters μ = 3.2 and σ = 1. What is the probability that the concentration exceeds 8 parts per million? (Table A.3, p735)
  - Solution Let X be the pollutant concentration

Since ln(X) has a normal distribution with $\mu = 3.2$ and $\sigma = 1$

$$P(X > 8) = 1 - P(X \le 8) = 1 - \phi[\frac{\ln(8) - 3.2}{1}]$$

$$= 1 - \phi[\frac{2.08 - 3.2}{1}] = 1 - \phi(-1.12) = 1 - 0.1314 = 0.8686$$

Here, we use the ϕ notation to denote the cumulative distribution function of the standard normal distribution.

Lognormal Distribution

- Example 6.23: The life, in thousands of miles, of a certain type of electronic control for locomotives (火車) has an approximate lognormal distribution with $\mu = 5.149$ and $\sigma = 0.737$. Find the 5th percentile of the life of such locomotive?
  - Solution

Lognormal Distribution

- Example 6.23: The life, in thousands of miles, of a certain type of electronic control for locomotives (火車) has an approximate <u>lognormal distribution</u> with μ = 5.149 and σ = 0.737. Find the 5th percentile of the life of such locomotive?
  - Solution

$$P(Z < z_1) = 0.05 \Rightarrow z_1 = -1.645$$

$$\because \ln(x) \text{ has a normal distribution with } \mu = 5.149 \text{ and } \sigma = 0.737$$

$$\therefore \frac{\ln(x) - 5.149}{0.737} = -1.645 \Rightarrow \ln(x) = 0.737(\text{-}1.645) + 5.149 = 3.937$$

$$\therefore x = 51.265$$

5% of the locomotives have lifetime less than 51.265 thousand miles.

Weibull Distribution

- Weibull distribution, introduced by the Swedish physicist Waloddi Weibull in 1939, has been used (like the gamma/exponential distribution) extensively in recent years to deal with the problems, e.g., a fuse may burn out, a steel column may buckle (彎曲), or a heat-sensing device may fail.
- Weibull Distribution: The continuous random variable X has a Weibull distribution with parameters α and β if its density function is given by

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$.

Gamma Distribution :

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$.

Weibull Distribution

- The mean and variance of the Weibull distribution are

$$\mu = \alpha^{-1/\beta}\Gamma(1+\tfrac{1}{\beta}) \text{ and } \sigma^2 = \alpha^{-2/\beta}\{\Gamma(1+\tfrac{2}{\beta}) - [\Gamma(1+\tfrac{1}{\beta})]^2\}.$$

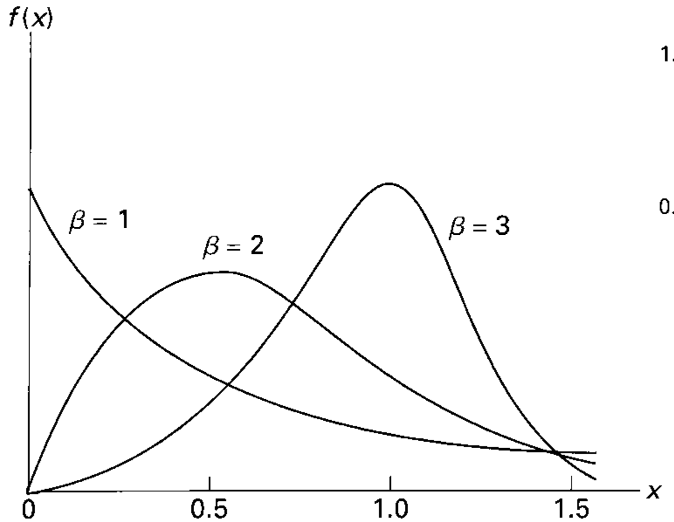

Figure 6.30 Weibull distributions ($\alpha = 1$).

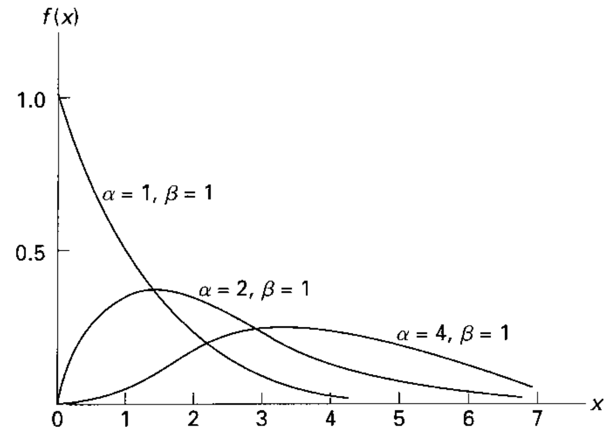

Figure 6.28 Gamma distributions.

Weibull Distribution

- Apply the Weibull distribution to reliability theory
  - Reliability: the probability that a component will function properly for at least a specified time under specified experimental conditions
  - If $f(t)$ is the Weibull distribution of the time of the component failure
  - If $R(t)$ is reliability of the component at time t, we may write

$$R(t) = P(T > t) = \int_t^{\infty} f(t)dt = 1 - \int_0^t f(t)dt = 1 - F(t)$$

The conditional probability that a component will fail in the interval

$$T = t \text{ to } T = t + \Delta t: \frac{F(t+\Delta t) - F(t)}{R(t)}$$

Dividing this ratio by Δt and taking the limit as $\Delta t \to 0$, we get the failure rate

$$Z(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t} \cdot \frac{1}{R(t)} = \frac{F'(t)}{R(t)} = \frac{f(t)}{R(t)} = \frac{f(t)}{1 - F(t)}$$

$$\because R(t) = 1 - F(t) \therefore R'(t) = -F'(t) = -f(t) \Rightarrow Z(t) = \frac{-R'(t)}{R(t)} = \frac{-d[\ln R(t)]}{dt}$$

$$\Rightarrow \ln R(t) = -\int Z(t)dt + \ln c \quad \text{or} \quad R(t) = ce^{-\int z(t)dt}$$

where c satisfies the initial assumption that $R(0) = 1$ or $F(0) = 1 - R(0) = 0$