1

Normal Distribution and Discrete Populations

138

138

Normal Distribution and Discrete Populations

- **Normal distribution** is often used as an approximation to the distribution of values in **discrete population**.
- In such situations, extra care should be taken to ensure that probabilities are computed in an accurate manner.

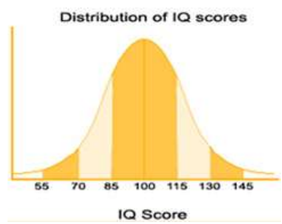
139

139

Example 4.19

- Intelligence Quotient (IQ) in a particular population (as measured by a standard test) is known to be approximately normally distributed with $\mu = 100$ and $\sigma = 15$.

- What is the probability that a randomly selected individual has an **IQ of at least 125?**



ระดับ	ไอคิว	ร้อยละ
อัจฉริยะ	>144	0.13
ปัญญาเลิศ	130-144	2.14
เหนือค่าเฉลี่ย	115-129	13.59
สูงกว่าค่าเฉลี่ย	100-114	34.13
ค่อนข้างต่ำ	85-99	34.13
ต่ำกว่าค่าเฉลี่ย	70-84	13.59
คาบเส้น	55-69	2.14
ต่ำ	<55	0.13

140

140

Example 4.19

Solution

- Letting X = the IQ of a randomly chosen person, we wish

$$P(X \geq 125).$$

- The temptation here is to standardize $X \geq 125$ as in previous examples.

- However, the IQ population distribution is actually discrete, since IQs are integer-valued.

141

141

EX 4.19

กำหนดให้ ตัวแปรสุ่ม X เป็น IQ ของบุคคลที่ถูกสุ่มเลือก

โจทย์ต้องการหา $P(X \geq 125) \rightarrow$ ทำ Standardize

$\mu = 100$

$\sigma = 15$

$$X \geq 125$$

$$\frac{X-\mu}{\sigma} \geq \frac{125-100}{15}$$

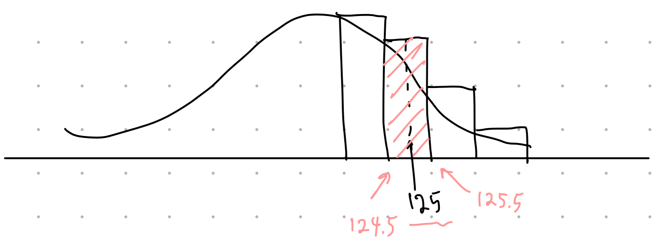
$$\Downarrow$$

$$P(Z \geq \underrightarrow{1.67})$$ เปิดตาราง

$$= 0.9525$$

เนื่องจาก IQ เป็น จน.เต็ม ⇒ การแจกแจงประชากรของ IQ เป็นแบบไม่ต่อเนื่อง (Discrete)

ดังนั้น Normal curve จึงเป็นการประมาณ Histogram จึงเป็นแบบไม่ต่อเนื่อง



เราต้องการหา พท. ใต้เส้นประมาณปกติ (Approximating)

$$X = 124.5$$

$$P(X \geq 125) \approx P(X \geq 124.5)$$

$$P\left(\frac{X-\mu}{\sigma} \geq \frac{125-100}{15}\right) \approx P\left(\frac{X-\mu}{\sigma} \geq \frac{124.5-100}{15}\right)$$

$$P(Z \geq 1.67) \approx P(Z \geq 1.6333\ldots)$$

$$0.0475 \approx 0.0516$$

$$1 - P(Z < 1.67) \approx 1 - P(Z \leq 1.63)$$

$$1 - 0.9525 \approx 1 - 0.9484$$

$$0.0475 \approx 0.0516$$ → ถูกต้องแม่นยำกว่า

Example 4.19

cont'd

- So normal curve is approximation to a discrete probability histogram, as pictured in Figure 4.24.

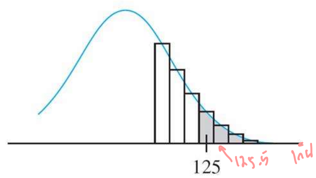


Figure 4.24 Normal approximation to a discrete distribution

- Rectangles of histogram are *centered* at integers, so IQs of at least 125 correspond to rectangles beginning at 124.5, as shaded in Figure 4.24.

- Thus we really want the area under the approximating normal curve to the right of 124.5.

142

142

Example 4.19

cont'd

$$Z = \frac{X - \mu}{\sigma} \qquad Z = \frac{124.5 - 100}{15} = \frac{24.5}{15} = 1.63333 \qquad Z = \frac{125 - 100}{15} = \frac{25}{15} = 1.66666$$

- Standardizing this value (124.5) gives $P(Z \geq 1.63) = 0.0516$, whereas standardizing 125 results in $P(Z \geq 1.67) = 0.0475$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
				[illegible]	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	[illegible]	[illegible]	[illegible]	[illegible]

- Difference is not great, but the answer 0.0516 is more accurate.
- Similarly, $P(X = 125)$ would be approximated by area between 124.5 and 125.5, since area under normal curve above the single value 125 is zero.

143

143

Example 4.19

cont'd

- Correction for discreteness of the underlying distribution in Example 19 is often called a **continuity correction.**

- It is useful in the following application of normal distribution to the computation of binomial probabilities.

144

2

Approximating Binomial Distribution

145

Approximating Binomial Distribution

- Recall that mean value and standard deviation of binomial random variable X are

$$\boxed{\mu_X = np}$$

$$\boxed{\sigma_X = \sqrt{npq}}$$

146

Approximating Binomial Distribution

Figure 4.25 displays a binomial probability histogram for binomial distribution with $n = 20$, $p = 0.6$, for which

$$\mu_X = np = 20(0.6) = 12 \quad \text{and}$$

$$\sigma_X = \sqrt{npq} = \sqrt{20(0.6)(0.4)} = 2.19$$

μ, σ

standard normal

$$Z = \frac{X - \mu}{\sigma}$$

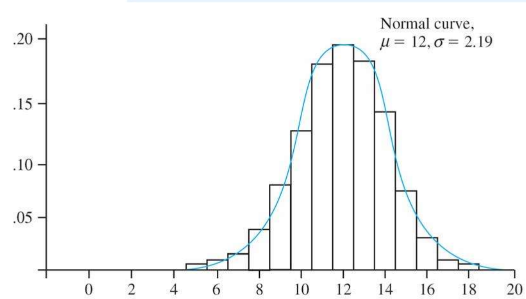


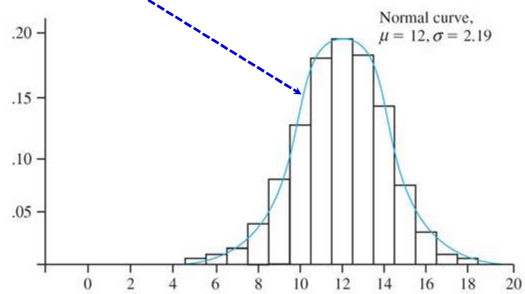
Figure 4.25 Binomial probability histogram for $n = 20$, $p = 0.6$ with normal approximation curve superimposed

147

Approximating Binomial Distribution

Normal curve with this μ and σ has been superimposed on probability histogram.

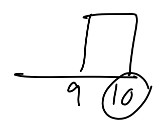
Although probability histogram is a bit skewed (because $p \neq 0.5$), normal curve gives a very good approximation, especially in the middle part of picture.



Area of any rectangle (probability of any particular X value) except those in the extreme tails can be accurately approximated by the corresponding normal curve area.

148

148

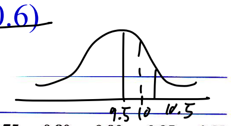


$P(X \leq 10) - P(X \leq 9)$

Approximating Binomial Distribution

For example, $P(X = 10) = B(10; 20, 0.6) - B(9; 20, 0.6)$

$= 0.245 - 0.128 = 0.117,$



n=20

		p														
		0.01	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
x	9	1.000	1.000	1.000	.997	.986	.952	.755	.412	.128	.017	.004	.001	.000	.000	.000
	10	1.000	1.000	1.000	.999	.996	.983	.872	.588	.245	.048	.014	.003	.000	.000	.000
	11	1.000	1.000	1.000	1.000	.999	.995	.943	.748	[illegible]	.113	.041	.010	.000	.000	.000

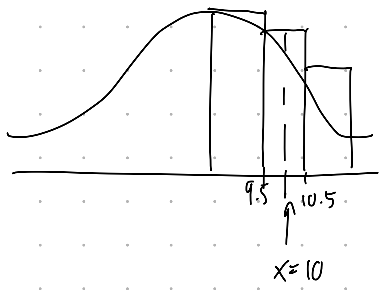
whereas area under the normal curve between 9.5 and 10.5 is

$P(-1.14 \leq Z \leq -0.68) = 0.1212.$ $\mu = 12$ $\sigma = 2.19$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170

149

149



$\mu = 12$
$\sigma = 2.19$

$$P(9.5 \le X \le 10.5) = P\left(\frac{9.5-12}{2.19} \le Z \le \frac{10.5-12}{2.19}\right) = P(-1.1415 \le Z \le -0.6849)$$

$$= P(Z \le -0.6849) - P(Z \le -1.1415)$$

$$= \Phi(-0.68) - \Phi(-1.14)$$

$$= 0.2483 - 0.1271$$

$$= 0.1212 \longrightarrow 0.117$$

[illegible]
Binomial

Approximating Binomial Distribution

- More generally, as long as binomial probability histogram is not too skewed, binomial probabilities can be well approximated by normal curve areas.

- It is then customary to say that X has approximately a normal distribution.

150

150

Approximating Binomial Distribution

Proposition

- Let X be binomial random variable based on n trials with success probability p.
- Then if binomial probability histogram is not too skewed, X has approximately normal distribution with $\mu = np \quad and \quad \sigma = \sqrt{npq}$

In particular, for x = possible value of X,

$$P(X \le x) = B(x, n, p) \approx \begin{pmatrix} \text{area under normal curve} \\ \text{to the left of } x + 0.5 \end{pmatrix} \qquad = \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right)$$

- In practice, approximation is adequate provided that both **$np \ge 10$** and **$nq \ge 10$**, since there is enough symmetry in underlying binomial distribution.

151

151

Example 4.20

- Suppose that 25% of all students at a large public university
ได้รับ ทุนการศึกษา
receive financial aid.

- Let X be the number of students in a random sample of size 50 n who receive financial aid, so that $p = 0.25$.

$\boxed{\mu_X = np}$ $np = 50(0.25)$ $\boxed{\sigma_X = \sqrt{npq}}$ $= \sqrt{50(0.25)(0.75)}$

Then $\mu = 12.5$ and $\sigma = 3.06$.

ตรวจสอบ
ว่าเข้ามาใกล้ป่าว

$np \geq 10$ $nq \geq 10$

- Since $np = 50(0.25) = 12.5 \geq 10$ and $nq = 50(0.75) = 37.5 \geq 10$, the approximation can safely be applied.

153

153

Example 4.20

$P(X \leq x) = B(x, n, p) \approx \left(\begin{array}{c}\textit{area under normal curve} \\ \textit{to the left of } x+0.5\end{array}\right) = \Phi\left(\frac{x+0.5-np}{\sqrt{npq}}\right)$ cont'd

- Probability that at most 10 students receive aid is

$$P(X \leq 10) = B(10;50,0.25) \approx \Phi\left(\frac{10+0.5-12.5}{3.06}\right) = \Phi(-0.65) = 0.2578$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451

- Similarly, probability that between 5 and 15 (inclusive) of selected students receive aid is

$$P(5 \leq X \leq 15) = B(15;50,0.25) - B(4;50,0.25)$$

$$\approx \Phi\left(\frac{15.5-12.5}{3.06}\right) - \Phi\left(\frac{4.5-12.5}{3.06}\right) = \Phi(0.98) - \Phi(-2.61) = 0.8320$$

- Exact probabilities are 0.2622 and 0.8348, respectively, so approximations are quite good.

154

154

Exponential Distribution

158

158

Exponential Distributions

The family of exponential distributions provides probability models that are very widely used in engineering and science disciplines.

ตัวอย่างเช่น

- ปัญหาในระบบแถวคอย – เวลาในการให้บริการและช่วงห่างระหว่างการเข้ามารับบริการของลูกค้า
- ความน่าจะเป็นที่ผลิตภัณฑ์ยังคงทำงานได้
- ความน่าจะเป็นที่ผลิตภัณฑ์จะเกิดการพังของอุปกรณ์หรือเครื่องจักรก่อนช่วงเวลาที่กำหนด
- อายุการใช้งานของอุปกรณ์

Definition

X is said to have an **exponential distribution** with parameter λ ($\lambda > 0$) if the pdf of *X* is

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & otherwise \end{cases} \qquad \textbf{(4.5)}$$

159

159

Exponential Distributions

- Some sources write the exponential pdf in the form $(1/\beta)e^{-x/\beta}$, so that $\beta = 1/\lambda$.
- Expected value of an exponentially distributed random variable X is

$$E(X) = \int_0^{\infty} x\lambda e^{-\lambda x}\, dx$$

- Obtaining this expected value necessitates doing an integration by parts.

The variance of X can be computed using the fact that

$$V(X) = E(X^2) - [E(X)]^2.$$

- The determination of $E(X^2)$ requires integrating by parts twice in succession.

160

160

Expected value

$$E(X) = \int_0^{\infty} x\lambda e^{-\lambda x}\, dx$$

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

$$E(X) = \lambda\int_0^{\infty} x e^{-\lambda x}\, dx$$

[illegible] $u = x \therefore du = dx$

$dv = e^{-\lambda x}dx \therefore v = \int_0^{\infty} e^{-\lambda x}dx$ $\quad\int e^{ax}dx = \frac{1}{a}e^{ax}$

$v = -\frac{1}{\lambda}e^{-\lambda x}$

Integrating by parts

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

$$\lambda\int_0^{\infty} x e^{-\lambda x}\,dx = \cancel{\lambda}\left[\left(x\cdot\frac{-1}{\cancel{\lambda}}e^{-\lambda x}\right)\Big|_0^{\infty} - \int_0^{\infty}\frac{-1}{\cancel{\lambda}}e^{-\lambda x}dx\right]$$

$$= -xe^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}dx$$

$$= -xe^{-\lambda x}\Big|_0^{\infty} + \left(\frac{1}{-\lambda}\right)e^{-\lambda x}\Big|_0^{\infty}$$

$$= (0-0) + \left(0 - \left(-\frac{1}{\lambda}\right)\right)$$

$$= \frac{1}{\lambda}$$

E(X) w.o Exponential

161

161

Variance — Exponential

$$Var[X] = E(X^2) - (E[X])^2 \qquad \int e^{au} = \frac{1}{a}\cdot e^{au}$$

$$E(X^2) = \int_0^\infty u^2 \cdot \lambda e^{-\lambda u}\,du$$

miunki $u = u^2 \quad \therefore du = 2u\,du$

$$dv = e^{-\lambda u}du \quad \therefore \quad v = \int_0^\infty e^{-\lambda u}du = -\frac{1}{\lambda}\cdot e^{-\lambda u}$$

Integrating by parts

$$\int u\,dv = uv - \int v\,du \qquad \text{By part}$$

$$\int_0^\infty u^2 \cdot \lambda e^{-\lambda u}du = \lambda \int_0^\infty u^2 \cdot e^{-\lambda u}du$$

$$= \cancel{\lambda}\left[\left(u^2 \cdot -\frac{1}{\cancel{\lambda}} e^{-\lambda u}\right)_0^\infty - \int_0^\infty -\frac{1}{\cancel{\lambda}}\cdot e^{-\lambda u}\cdot 2u\,du\right]$$

$$= (0-0) + \int_0^\infty 2u e^{-\lambda u}du \qquad \text{By part}$$

$$= 2\int_0^\infty u e^{-\lambda u}du \qquad \int e^{au} = \frac{1}{a}\cdot e^{au}$$

Qu: $u = u \quad \therefore du = du$

$$dv = e^{-\lambda u}du \quad \therefore \quad v = \int_0^\infty e^{-\lambda u}du = -\frac{1}{\lambda}e^{-\lambda u}$$

$$= 2\left[\left(u \cdot -\frac{1}{\lambda}e^{-\lambda u}\right)_0^\infty - \int_0^\infty -\frac{1}{\lambda}e^{-\lambda u}du\right]$$

$$= 2\left[\left(u \cdot -\frac{1}{\lambda}e^{-\lambda u}\right)_0^\infty + \int_0^\infty \frac{1}{\lambda}e^{-\lambda u}du\right]$$

$$= 2\left[(0-0) + \frac{1}{\lambda}\int_0^\infty e^{-\lambda u}du\right] \qquad \int e^{au} = \frac{1}{a}e^{au}$$

$$= \frac{2}{\lambda}\left(-\frac{1}{\lambda}\cdot e^{-\lambda u}\right)_0^\infty$$

$$= \frac{2}{\lambda^2}\left(-e^{-\lambda u}\right)_0^\infty$$

$$= \frac{2}{\lambda^2}\left(-0-(-1)\right)$$

$$= \frac{2}{\lambda^2}$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

162

Exponential Distributions

The results of these integrations are as follows:

$$\mu = \frac{1}{\lambda} \qquad \sigma^2 = \frac{1}{\lambda^2}$$

Both the mean and standard deviation of the exponential distribution equal $1/\lambda$.

Graphs of several exponential pdf's are illustrated in Figure 4.26.

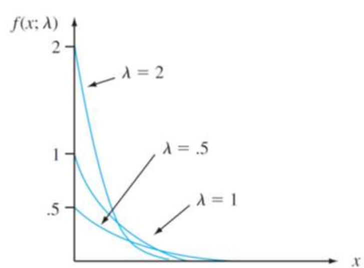


Figure 4.26 Exponential density curves

163

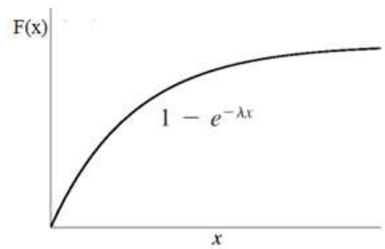
Exponential Distributions

The exponential pdf is easily integrated to obtain the cdf.

$$F(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

$\lambda = 0.1018$

$1 - e^{-\lambda x}$



164

164

$F(x;\lambda)$

pdf; $f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$

$$F(x;\lambda) = P(X \leq x) = \int_{-\infty}^{x} f(x;\lambda)\, dx$$

$$= \int_{0}^{x} \lambda e^{-\lambda x}\, dx$$

$$= \lambda \int_{0}^{x} e^{-\lambda x}\, dx \qquad \int e^{ax} = \frac{1}{a}\cdot e^{ax}$$

$$= \cancel{\lambda}\left(\frac{1}{\cancel{-\lambda}}\cdot e^{-\lambda x}\Big|_{0}^{x}\right)$$

$$= \left(-e^{-\lambda x}\Big|_{0}^{x}\right)$$

$$= -e^{-\lambda x} - (-e^{-\lambda(0)}) \quad 1$$

$$= -e^{-\lambda x} + 1$$

$$= \boxed{1 - e^{-\lambda x}}$$

$\int e^{-\lambda x}\, dx$

ให้ $u = -\lambda x$

$du = \frac{du}{dx}\cdot dx = \frac{d(-\lambda x)}{dx}\, dx$

$= -\lambda\, dx$

$= dx \cdot \frac{1}{-\lambda}\, du$

$\int e^{-\lambda x}\, dx = \frac{-1}{\lambda}\int e^{u}\, du$

$= -\frac{1}{\lambda} e^{u}$

$= -\frac{1}{\lambda} e^{-\lambda x}$

165

165

Example 4.21

MPa : MegaPascals

The article "Probabilistic Fatigue Evaluation of Riveted Railway Bridges" (*J. of Bridge Engr.*, 2008: 237–244) suggested the exponential distribution with mean value 6 MPa as a model for the distribution of stress range in certain bridge connections.

E(X)

Let's assume that this is in fact the true model. Then

$E(X) = 1/\lambda = 6$ implies that $\lambda = 0.1667$.

166

166

Example 4.21

cont'd

The probability that stress range is at most 10 MPa is

F(x ; λ)

$$P(X \le 10) = F(10\ ;\ 0.1667)$$

$$= 1 - e^{-(0.1667)(10)}$$

$$F(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

$$= 1 - 0.189$$

$$= 0.811$$

167

167

Example 4.21

$$F(x; \lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

cont'd

The probability that stress range is between 5 and 10 MPa is

$$P(5 \leq X \leq 10) = F(10; 0.1667) - F(5; 0.1667)$$

$$= (1 - e^{-0.1667(10)}) - (1 - e^{-0.1667(5)})$$

$$= (1 - e^{-1.667}) - (1 - e^{-0.8335})$$

$$= 0.246$$

Exponential Distributions

- Exponential distribution is frequently used as a model for distribution of times between occurrence of successive events, such as
  - customers arriving at a service facility or
  - calls coming in to a switchboard.

170

170

Exponential Distributions

Proposition

- Suppose that the number of events occurring in any time interval of length t has a Poisson distribution with parameter αt (where α, rate of event process, is expected number of events occurring in 1 unit of time) and that numbers of occurrences in nonoverlapping intervals are independent of one another.
- Then distribution of elapsed time between occurrence of two successive events is exponential with parameter $\lambda = \alpha$.

171

171

Exponential
X
Poisson

Exponential Distributions

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Although a complete proof is beyond the scope of the text, the result is easily verified for the time X_1 until the first event occurs:

$$P(X_1 \le t) = 1 - P(X_1 > t) = 1 - P\,[\text{no events in } (0, t)]$$

$$= 1 - \frac{e^{-\alpha t} \cdot (\alpha t)^0}{0!} = 1 - e^{-\alpha t}$$

which is exactly the cdf of the exponential distribution.

$$F(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

172

172

Example 4.22

○ Suppose that calls are received at a 24-hour "suicide hotline" according to a Poisson process with rate $\alpha = 0.5$ call per day.

○ Then the number of days X between successive calls has an exponential distribution with parameter value 0.5, so the probability that more than 2 days elapse between calls is

$$\begin{aligned} P(X > 2) &= 1 - P(X \le 2) \\ &= 1 - F(2; 0.5) \\ &= e^{-(0.5)(2)} \\ &= 0.368 \end{aligned}$$

$$F(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

$$\mu = \frac{1}{\lambda}$$

The expected time between successive calls is $1/0.5 = 2$ days.

173

173

Exponential Distributions

Another important application of the exponential distribution is to model the distribution of component lifetime.

A partial reason for the popularity of such applications is the **"memoryless" property** of the exponential distribution.

Suppose component lifetime is exponentially distributed with parameter λ.

174

174

Exponential Distributions

After putting the component into service, we leave for a period of t_0 hours and then return to find the component still working; what now is the probability that it lasts at least an additional t hours?

In symbols, we wish $P(X \geq t + t_0 \mid X \geq t_0)$.

By the definition of conditional probability,

$$P(X \geq t + t_0 \mid X \geq t_0) = \frac{P[(X \geq t + t_0) \cap (X \geq t_0)]}{P(X \geq t_0)}$$

175

175

Exponential Distributions

$$F(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

But the event $X \geq t_0$ in the numerator is redundant, since both events can occur if $X \geq t + t_0$ and only if. Therefore,

$$P(X \geq t + t_0 \mid X \geq t_0) = \frac{P(X \geq t + t_0)}{P(X \geq t_0)} = \frac{1 - F(t + t_0; \lambda)}{1 - F(t_0; \lambda)} = e^{-\lambda t}$$

This conditional probability is identical to the original probability $P(X \geq t)$ that the component lasted t hours.

- Thus *distribution of additional lifetime is exactly the same as the original distribution of lifetime,* so at each point in time the component shows no effect of wear.

- In other words, the distribution of remaining lifetime is independent of current age.

176

177