Continuous Probability Distributions

[Navidi sections 4.5-4.8 and 4.10; Devore sections 4.3-4.5]

Chapter 6 introduced the concept of probability density functions for continuous random quantities. The only standard distribution introduced then is the **continuous uniform distribution**, *U*(*a*, *b*), whose p.d.f. is

$$f(x) = \frac{1}{b-a} \qquad (a \le x \le b)$$

with

$$\mathrm{E}[X] = \frac{a+b}{2} \quad \text{and} \quad \mathrm{V}[X] = \frac{(b-a)^2}{12}$$

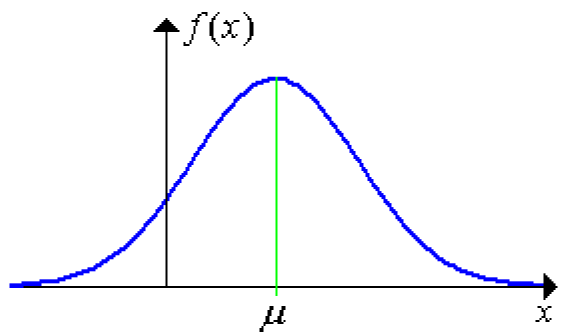
This chapter will introduce some more standard probability density functions.

The Normal Distribution

The Gaussian or normal probability distribution is the single most important probability distribution. It was first described by Abraham de Moivre in 1733 but bears the name of Karl Friedrich Gauss, who arrived at this distribution in 1809 when examining the distribution of errors in the measurement of the diameters of lunar craters. It arises naturally in many other situations (especially the Central Limit Theorem).

If a continuous random quantity X has a normal distribution with population mean μ and population variance σ^2, then its probability density function is

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$



which is positive for all x.
It is symmetric about $x = \mu$.

The cumulative distribution function for the normal distribution is

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

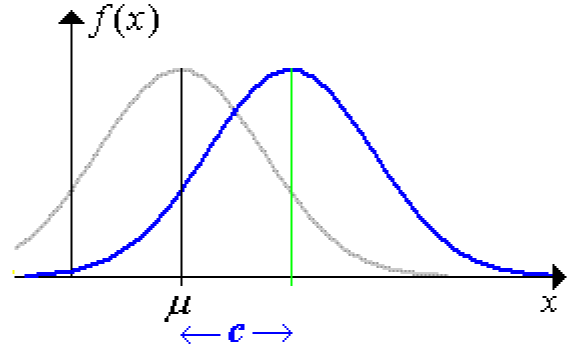
which cannot be evaluated exactly in closed form except for certain special choices for x.
It can be shown that $\lim_{x \to \infty} F(x) = 1$, as required by the condition for coherent probabilities.

Notation: $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$

$\mu = \mathrm{E}[X] =$ population mean

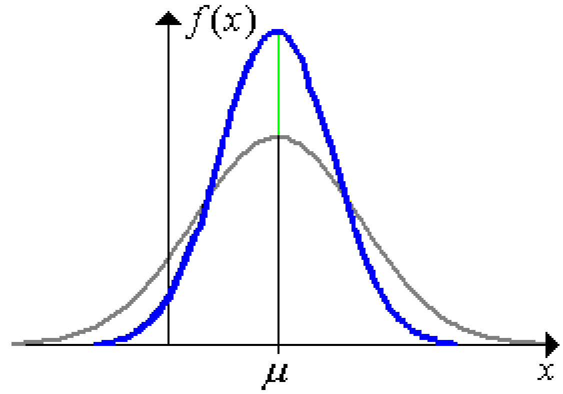
$\sigma^2 = \mathrm{V}[X] =$ population variance

Adding a constant c to X:



Influence of the variance σ^2 on the shape of the normal probability curve:

Low σ^2 High σ^2



Because there is no closed algebraic form for the *c.d.f.* $F(x)$, the values are tabulated for one special choice of mean and variance: $\mu = 0,\ \sigma^2 = 1$.

Notation:

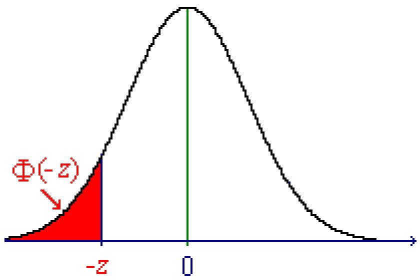
$Z \sim \mathrm{N}\ (0,\ 1)$ is the **standard normal distribution**,

with *p.d.f.* = $\phi(z)$ and *c.d.f.* = $\Phi(z)$.

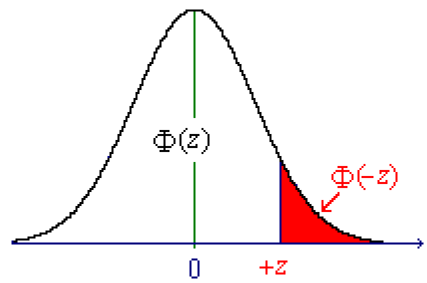
Conversion from $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$ to $Z \sim \mathrm{N}\ (0,\ 1)$ requires a linear shift of μ and a change of scale by a factor of σ.

$$F(x) = \mathrm{P}[X \le x] =$$

Thus $\mathrm{P}[x_1 < X < x_2] = \Phi(z_2) - \Phi(z_1)$ where $z = \dfrac{x - \mu}{\sigma}$.



Symmetry $\Rightarrow$



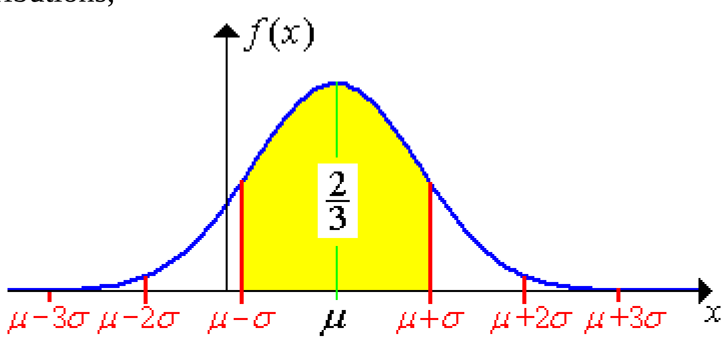
$$\boxed{\Phi(-z) \quad = \quad 1 - \Phi(+z)}$$

A table of values of the standard normal distribution is on the inside front cover of the textbooks (Navidi, Table A.2; Devore Table A.3).
A more precise table is available as an Excel spreadsheet file, on the course web site, at
"`www.engr.mun.ca/~ggeorge/4421/demos/zTables.xls`".
and also in Chapter 17 of these notes.

Among the useful properties of the Normal distribution is that, unlike other probability distributions, any linear combination of normally distributed random quantities is also normally distributed. For an uncorrelated set of n random quantities $\{ X_i \}$,

if $X_i \sim \text{N}\left(\mu_i, \sigma_i^{\,2}\right)$ then $Y = \sum_{i=1}^{n} a_i X_i \sim \text{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^{\,2} \sigma_i^{\,2}\right)$

For all normal distributions,

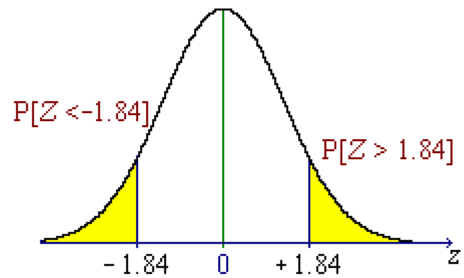


$$\text{P}\left[\,|X-\mu| > 1\sigma\,\right] \approx \frac{1}{3} \qquad\qquad \text{P}\left[\,|X-\mu| > 2\sigma\,\right] \approx \frac{1}{20}$$

$$\text{P}\left[\,|X-\mu| > 3\sigma\,\right] < \frac{1}{300}$$

Example 10.01:

The weights of boxes of nails are known to be normally distributed to an excellent approximation, with mean 454 grammes and standard deviation 25 grammes. What proportion of boxes weighs more than 500 grammes?

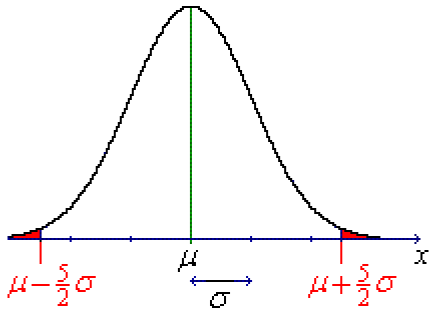


Part of Table A.2 (also page 17-03):

z	**.00**	**.01**	**.02**	**.03**	**.04**	**...**
⋮						
–1.9	0.02872	0.02807	0.02743	0.02680	0.02619	**...**
–1.8	0.03593	0.03515	0.03438	0.03362	0.03288	**...**
–1.7	0.04457	0.04363	0.04272	0.04182	0.04093	**...**
–1.6	0.05480	0.05370	0.05262	0.05155	0.05050	**...**
–1.5	0.06681	0.06552	0.06426	0.06301	0.06178	**...**
⋮						

Example 10.02

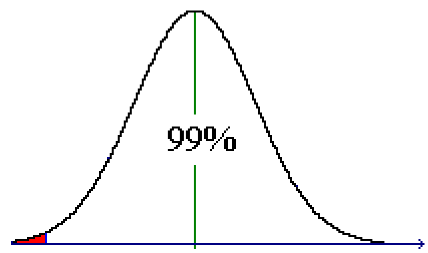
Find the probability that a normally distributed random quantity is more than two and a half standard deviations away from its mean in either direction.



$$\mathrm{P}\left[\,|X-\mu| > \left(2\tfrac{1}{2}\right)\sigma\,\right] = \mathrm{P}\left[\left(X-\mu > \frac{5}{2}\sigma\right) \cup \left(X-\mu < -\frac{5}{2}\sigma\right)\right]$$

Example 10.03

The strength of a set of steel bars is known to be normally distributed with a population mean of 5 kN and a population variance of $(50 \text{ N})^2$. A client requires that at least 99% of all of these bars be stronger than 4,900 N. Has this requirement been met?



Example 10.04

Given that the random quantity X is normally distributed, find c such that

$$\mathrm{P}[\, X > c \,] = 1\% \,.$$

(That is, find the 99th percentile.)

$$\mathrm{P}\left[X > c\right] \;=\; \mathrm{P}\left[Z > \frac{c-\mu}{\sigma}\right] \;=\; .01 \,.$$

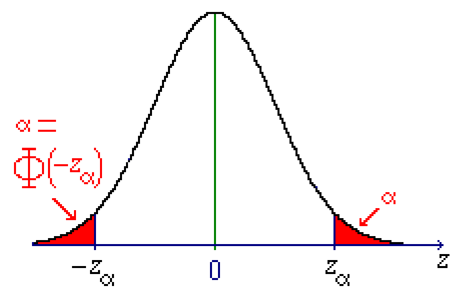
Notation: z_α = the $(1 - \alpha)\times 100$th percentile of the standard normal distribution.

$$\mathrm{P}\left[Z > z_\alpha\right] = \alpha = \Phi\left(-z_\alpha\right)$$

Here, $\alpha = .01$.

$\Phi\left(-z_{.01}\right) = .01$

To find $z_{.01}$ we need to look in table A.2 for the value $\Phi = .01000$.



$$\Phi(-2.32) \;=\; .01017$$
$$\Phi(-2.33) \;=\; .00990$$

$\Rightarrow \quad z_{.01} = 2.33$ (to 2 d.p.) [as in the previous example]

Using *linear interpolation* (which will *not* be required in tests or the exam):

.01000 is 17/27 of the way from $\Phi(-2.32)$ to $\Phi(-2.33)$.

Thus $-z_{.01} \approx -2.32 + \dfrac{17}{27}\times\left(-0.01\right) \approx -2.326$.

[The true value is –2.326, correct to three decimal places].

$$\frac{c-\mu}{\sigma} = z_{.01} \approx 2.326 \quad\Rightarrow\quad c = \mu + z_{.01}\sigma \approx \mu + 2.326\sigma \,.$$

Therefore the 99$^{\text{th}}$ percentile of any normal distribution is 2.326 standard deviations above the mean.

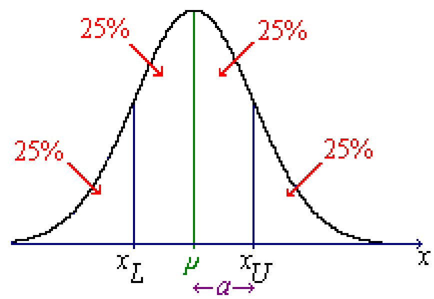
Note: By symmetry, the fiftieth percentile ($z_{.50} = \tilde{\mu} =$ the median) is at $z = 0$.

$$\Phi\left(0\right) = .50000 \quad\Rightarrow\quad \tilde{\mu} = \mu + 0\sigma \,.$$

Example 10.05

Find the quartiles for any normal distribution.

$$X \sim \mathrm{N}\left(\mu, \sigma^2\right)$$



By symmetry,

$x_L = \mu - a$ and

$x_U = \mu + a$ where

a = SIQR (the semi-interquartile range).

$$F\left(x_U\right) = .75 \quad \Rightarrow \quad \Phi\left(\frac{x_U - \mu}{\sigma}\right) = .75000$$

But $\Phi\,(0.67) = .74857$ and $\Phi\,(0.68) = .75175$.

$\Rightarrow$ (to 2 s.f.) **SIQR = <u>0.67 σ</u>**

Linear interpolation:

$.75000 \;=\; \Phi\Bigg($

Example 10.06

Pistons used in a certain type of engine have diameters that are known to follow a normal distribution with population mean 22.40 cm and population standard deviation 0.03 cm. Cylinders for this type of engine have diameters that are also distributed normally, with mean 22.50 cm and standard deviation 0.04 cm.

What is the probability that a randomly chosen piston will fit inside a randomly chosen cylinder?

Let X_P = piston diameter $\sim \text{N}\left(22.40, (0.03)^2\right)$

and X_C = cylinder diameter $\sim \text{N}\left(22.50, (0.04)^2\right)$

Note that the difference of two independent normally distributed random quantities also follows a normal distribution.

The random selection of piston and cylinder $\Rightarrow$ X_P and X_C are <u>independent</u>.

The chosen piston fits inside the chosen cylinder iff $X_P < X_C$.

The normal distribution is symmetric, with over 99% of all values lying within three standard deviations of the mean. The inner fences of any boxplot (beyond which any values are outliers) are 3×(semi-interquartile range) beyond the quartiles. For a normal distribution, the quartiles are at $\mu \pm 0.6745\sigma$ (from Example 10.05 above).
The inner fences are at $\mu \pm 4 \times 0.6745\sigma = \mu \pm 2.698\sigma$.

For any normal distribution,

$$\mathrm{P}\left[|X-\mu| > 2.698\sigma\right] = 2\mathrm{P}\left[X-\mu < -2.698\sigma\right] = 2\mathrm{P}\left[\frac{X-\mu}{\sigma} < -2.698\right]$$

$$= 2\Phi(-2.698) \approx .00698$$

Therefore less than 0.7% of all values in a normal distribution are outliers.

The outer fences are at $\mu \pm 7 \times 0.6745\sigma = \mu \pm 4.7215\sigma$
and $\mathrm{P}\left[|X-\mu| > 4.7215\sigma\right] = 2\,\Phi(-4.7215) \approx .000\,002\,34$
These values of the standard normal c.d.f. $\Phi(z)$ were found from the Excel file
"`www.engr.mun.ca/~ggeorge/4421/demos/zCalculator.xls`".

Extreme outliers are therefore very rare for normal distributions, occurring by chance about once per half-million values.

If a sample shows only positive values with a very strong positive skew and several extreme outliers on the positive side only, then one of the candidates for the probability distribution is the lognormal distribution.

The Lognormal Distribution (***not*** examinable; here for future reference only)

If $X \sim \mathrm{N}(\mu, \sigma^2)$, then the random quantity $Y = e^X$ has the lognormal distribution with parameters μ and σ^2, probability density function

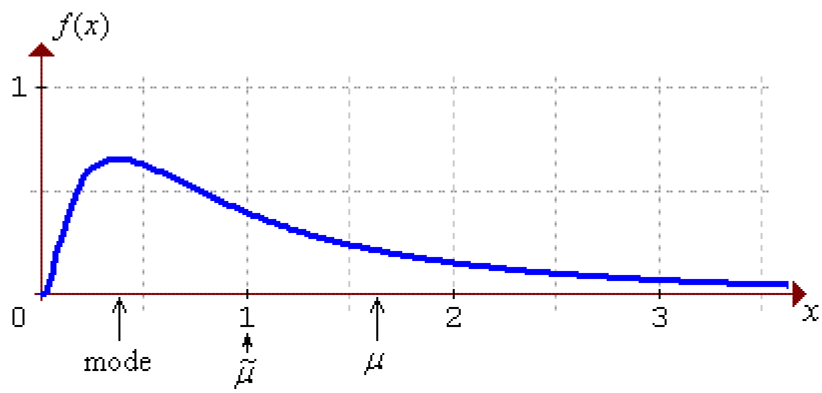
$$\boxed{f(y) = \frac{1}{\sigma y\sqrt{2\pi}}\, e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}} \qquad (y>0)}$$

and

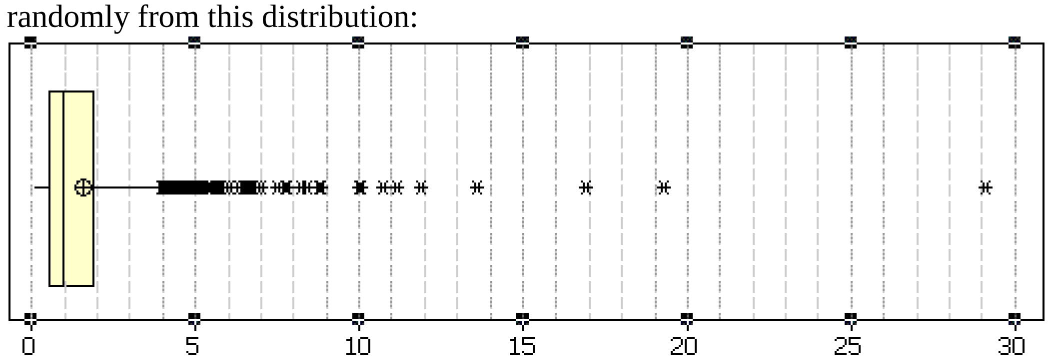
$$\boxed{\mathrm{E}[Y] = e^{\mu+\sigma^2/2}} \quad \text{and} \quad \boxed{\mathrm{V}[Y] = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}}$$

Conversely, if the random quantity X is lognormal with parameters μ and σ^2, then $\ln X$ is $\mathrm{N}(\mu, \sigma^2)$.

Graph of the standard lognormal probability density function $(\mu = 0,\ \sigma^2 = 1)$:



The mean value is $e^{\mu+\sigma^2/2} = e^{0+1/2} \approx 1.65$, which is far beyond the mode of ≈ 0.4.
The variance is $e^2 - e^1 \approx 4.67$. The extreme skew is obvious in a boxplot of 1000 values drawn randomly from this distribution:



Example 10.07

Find the median and the 80th percentile for the lognormal distribution $(\mu = 0,\ \sigma^2 = 1)$.

If X is lognormal $(\mu = 0,\ \sigma^2 = 1)$, then $\ln X \sim \text{N}(0, 1)$.

$$\text{P}[X < \tilde{\mu}] = \frac{1}{2} \quad \Rightarrow \quad \text{P}[\ln X < \ln \tilde{\mu}] = \frac{1}{2} \quad \Rightarrow \quad \text{P}\left[Z < \frac{\ln \tilde{\mu} - 0}{\sqrt{1}}\right] = \frac{1}{2}$$

$$\Rightarrow \quad \Phi(\ln \tilde{\mu}) = \frac{1}{2} = \Phi(0) \qquad \Rightarrow \quad \ln \tilde{\mu} = 0 \qquad \Rightarrow \quad \tilde{\mu} = \underline{\underline{\mathbf{1}}}$$

By a similar argument,

$$\text{P}[X < x_{.80}] = .80 \quad \Rightarrow \quad \Phi(\ln(x_{.80})) = .80000 \approx \Phi(0.84162)$$

$$\Rightarrow \quad \ln(x_{.80}) \approx 0.84162 \qquad \Rightarrow \quad x_{.80} \approx \underline{\underline{\mathbf{2.320}}}$$

Example 10.08

The time T to next failure of a certain process is known to be lognormally distributed. A random sample of seven runs of the process has lifetimes (in days) of

2.61114, 2.88284, 2.59738, 1.83647, 1.59441, 2.34576, 2.80628

(a) Estimate the parameters of the lognormal distribution
(b) Estimate the mean lifetime of the process.

(a) If T is lognormal, then $X = \ln T$ is normal.
Taking natural logarithms of the data:

{ 0.95979, 1.05878, 0.95450, 0.60784, 0.46651, 0.85261, 1.03186 }

$n = 7$, $\sum x = 5.9319$, $\sum x^2 = 5.3321$

$$\Rightarrow \quad \bar{x} = \frac{5.9319}{7} \approx 0.8474, \quad s^2 = \frac{7 \times 5.3321 - (5.9319)^2}{7 \times 6} = \frac{2.1372\ldots}{42} \approx 0.050887$$

$$\Rightarrow \quad s \approx 0.2256$$

Therefore the estimates of the parameters of the lognormal distribution for T are $\hat{\mu} \approx 0.8474$ and $\hat{\sigma} \approx 0.2256$.

(b) The estimate of the mean lifetime is

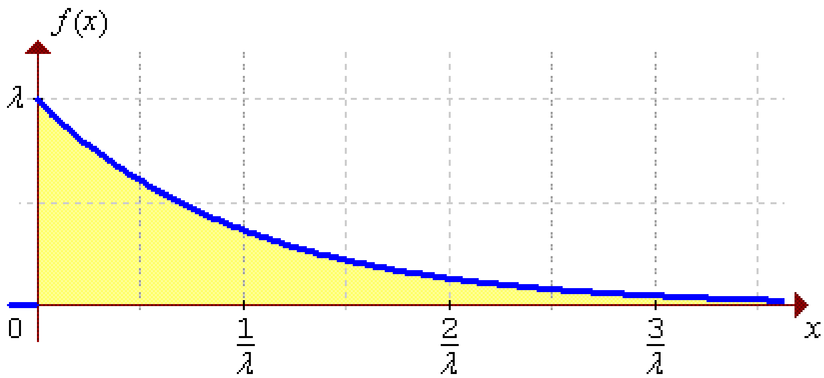
$$e^{\hat{\mu} + \hat{\sigma}^2/2} = e^{0.8474 + (0.0508\ldots)/2} \approx 2.39 \text{ days.}$$

The Exponential Distribution

This continuous probability distribution often arises in the consideration of lifetimes or waiting times and is a close relative of the discrete Poisson probability distribution. Like the lognormal distribution, the exponential distribution has a strong positive skew.

The probability density function is

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



The cumulative distribution function is

$$F(x) = \mathrm{P}[X \le x] = \int_{-\infty}^{x} f(t)\, dt$$

$$\Rightarrow \quad \mathrm{P}[X > x] =$$

Also $\mu = \mathrm{E}[X] =$ and $\sigma = \mu$

Example 10.09

The random quantity X follows an exponential distribution with parameter $\lambda = 0.25$.
Find μ , σ and P[$X > 4$] .

μ =

$\text{P}\left[X > 4\right]$ =

Note: For *any* exponential distribution, P[$X > \mu$] ≈ .368 .

Example 10.10

The arrival time T of the next customer follows an exponential distribution with a mean waiting time of five minutes.

(a) Find the probability that the next customer arrives within the next ten minutes.

λ =

P[$T \leq 10$] = $F(10)$ =

Relationship with the Poisson Distribution

If T is the waiting time until the next occurrence of an event and if T follows an exponential distribution with mean waiting time μ , then the parameter $\lambda = \dfrac{1}{\mu}$ is the average number of events per unit time.

If N is the number of events in a time interval τ , then N follows a Poisson distribution with the parameter $\lambda\tau$, (the number of events expected every τ seconds).

Example 10.10 (continued)

The arrival time T of the next customer follows an exponential distribution with a mean waiting time of five minutes.

(b) Find the probability that less than two customers arrive in the next ten minutes.

Let N = number of customers who arrive in the next ten minutes, then

$\lambda = \frac{1}{\mu} = \frac{1}{5} = .2$ per min. = 2 per (10 min.)

$\Rightarrow$ N follows a Poisson distribution with parameter $10\lambda = \mu = 2$

[We expect, on average, 2 customers every 10 minutes]

$\mathrm{P}[N < 2] = \mathrm{P}[N \le 1] =$

Note:

$$\mathrm{P}[X > \mu + 2\sigma] = e^{-\lambda(\mu+2\sigma)} = e^{-\lambda((1/\lambda)+(2/\lambda))} = e^{-3} \approx .049\,787$$

Therefore **P[$X > \mu + 2\sigma$] ≈ 5.0% for <u>all</u> exponential distributions.**

Also $\mu - \sigma = \frac{1}{\lambda} - \frac{1}{\lambda} = 0 \Rightarrow \mathrm{P}[X < \mu - \sigma] = 0 = \mathrm{P}[X < \mu - 2\sigma]$

Therefore **P[| $X - \mu$ | > 2σ] ≈ 5.0%** , a result similar to the normal distribution, except that *all* of the probability is in the upper tail only.

It can be shown that the sum of n independent random quantities, all of which follow an exponential distribution with the same parameter λ, is an **Erlang distribution** with parameters n and λ, $\Gamma(n,\lambda)$. Its probability density function, mean, variance and cumulative distribution function are

$$\boxed{f(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} \quad (t>0)}, \quad \boxed{\mu = \frac{n}{\lambda}}, \quad \boxed{\sigma^2 = \frac{n}{\lambda^2}} \quad \text{and}$$

$$\boxed{F(t) = \mathrm{P}[T \le t] = \begin{cases} 0 & (t \le 0) \\ 1 - \displaystyle\sum_{j=0}^{n-1} \frac{(\lambda t)^j}{j!} e^{-\lambda t} & (t > 0) \end{cases}}$$

The exponential distribution is clearly a special case: $\Gamma(1,\lambda)$.

Example 10.10 (continued)

The arrival time T of the next customer follows an exponential distribution with a mean waiting time of five minutes.

(c) Find the probability that the next two customers arrive in less than 20 minutes.

T follows an exponential distribution with $\mu = 5$ min $\Rightarrow$ $\lambda = 0.2 /$ min.
Let T_2 = waiting time for the second customer, then $T_2 \sim \Gamma(2, 0.2)$

$$\mathrm{P}[T_2 < 20] = F(20) =$$

The generalization of the Erlang distribution is the **gamma distribution** $\Gamma(r,\lambda)$, where r is a positive real number. The p.d.f. is

$$\boxed{f(t) = \frac{\lambda^r t^{r-1}}{\Gamma(r)} e^{-\lambda t} \quad (t>0)}, \quad \text{where the gamma function is} \quad \boxed{\Gamma(r) = \int_0^\infty t^{r-1} e^{-t} \, dt}$$

For reference purposes only, here are some other continuous probability distributions (Devore section 4.5; Navidi section 4.8):

Weibull distribution (parameters α and β)**:**

$$\boxed{f(x;\alpha,\beta) \;=\; \alpha\beta^{\alpha}\, x^{\alpha-1} e^{-(\beta x)^{\alpha}}\,, \qquad (x>0)}$$

The **exponential distribution** is a special case of

- the Weibull distribution when $\alpha = 1$. $(\lambda = \beta)$.
- the gamma distribution when $r = 1$.

However, neither of the Weibull and gamma distributions is a subset of the other.

Another special case of the gamma distribution, with $r = \nu/2$ and $\lambda = 1/2$ (where ν = a natural number = "degrees of freedom") is the **chi-squared distribution:**

$$\boxed{f(x;\nu) \;=\; \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\, x^{(\nu/2)-1}\, e^{-x/2}\,, \qquad (x>0)}$$

The chi-squared distribution will occur later, in Chapter 14.

Another distribution employed in many applications is the **beta distribution:**

$$f(x;\alpha,\beta,A,B) \;=\; \frac{1}{B-A}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\left(\frac{x-A}{B-A}\right)^{\alpha-1}\left(\frac{B-x}{B-A}\right)^{\beta-1}, \qquad A \le x \le B\,.$$

It can model completion times for projects, between an extreme optimistic time of A and an extreme pessimistic time of B.

Other *p.d.f.*s can be found in any good textbook on probability and statistics.

Probability Plots

Given a random sample, a probability plot can determine whether the sample is consistent with some specified population. This topic will be explored in a Minitab tutorial only.

[End of Chapter 10]

[Space for Additional Notes]