Chapter 3. Continuous Distributions

- ♣ Continuous-Type Data
- ♣ Exploratory Data Analysis
- ♣ Random Variables of The Continuous Type
- ♣ Uniform, Exponential, Gamma, Chi-Square, Normal, Beta, Student-t, and F Distributions
- ♣ Gamma, Exponential, and Chi-Square Distributions
- ♣ The Normal Distribution
- ♣ The Moment-Generating Functions
- ♣ Exercises for $\Gamma(\alpha, \theta), Expoenetial, \chi^2(r), N(\mu, \sigma^2)$

□ *Continuous-Type Data* Many experiments or observations of random phenomena do not have integers as outcomes, but instead are the measurements selected from an interval of numbers. For example, you could find the length of time that it takes when waiting in line to buy train tickets. Conceptually, if the measurements could come from an interval of possible outcomes, we call them *data from a continuous-type population* or *continuous-type data.*

Example One characteristic of a car's storage console that is checked by the manufacturer is the time in seconds that it takes for the lower storage compartment door to open completely. A random sample of size $n = 5$ yielded the following times:

$$1.1 \quad 0.9 \quad 1.4 \quad 1.1 \quad 1.0$$

Find the sample mean $\bar{x}$ and sample variance s^2.

Exploratory Data Analysis (1/2)

To explore the other characteristics of an unknown distribution, we need to take a sample of n observations, $x_1, x_2, \cdots, x_n$, from that distribution and often need to order them from the smallest to the largest. One convenient way of doing this is to use a stem-and-leaf display started by John W. Tukey (1977).

♢ stem-and-leaf display

♢ order statistics (of the sample)

♢ 25th *percentile*, 0.25 *quantile*, 1st *quartile*

♢ minimum, mean, median, maximum, range

♢ 1st quartile, 2nd quartile (median), 3rd quartile (q_3)

♢ five-number summary (minimum, q_1, q_2, q_3, maximum)

♢ box-and-whisker diagram, outliers

□ *Graphical Comparisons of Data Sets*

♢ A plot of Taiwan Stock Exchange Indices

♢ A plot of Dow Jones Industrial Indices

♢ A plot of Nasdaq Indices

Exploratory Data Analysis (2/2)

(1) There is a list of scores on a statistics examination provided by Professor John W. Tukey (1977).

93	77	67	72	52	83	66	84	59	63
75	97	84	73	81	42	61	51	91	87
34	54	71	47	79	70	65	57	90	83
58	69	82	76	71	60	38	81	74	69
68	76	85	58	45	73	75	42	93	65

(a) List the order statistics of these 50 scores.

(b) Find sample mean and variance for these scores.

(c) Find the 25th, 75th percentiles, and the median.

(d) Draw a box-and-whisker diagram, including a five-number summary.

(e) How may students were flunk?

(f) How may students receiving scores over 90, inclusive?

(g) Are there outliers? Explain it.

(h) Find the tabulation of these 50 scores.

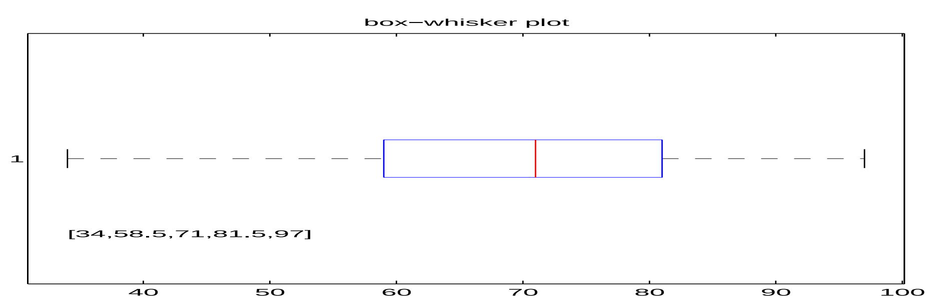
(i) Find the histogram of relative frequency.

(j) Find a (ordered) stem-leaf display of these scores.

(a) The order statistics of data in **(1)** are shown below.

34	38	42	42	45	47	51	52	54	57
58	58	59	60	61	63	65	65	66	67
68	69	69	70	71	71	72	73	73	74
75	75	76	76	77	79	81	81	82	83
83	84	84	85	87	90	91	93	93	97

(d) Box-and-Whisker Plot.



Random Variables of The Continuous Type

♣ Random variables whose space are intervals or a union of intervals are said to be of the *continuous types*. The *p.d.f.* of a r.v. X of continuous type is an integrable function $f(x)$ satisfying

(a) $f(x) > 0,\ x \in R$

(b) $\int_R f(x)dx = 1$

(c) The probability of the event $X \in A$ is $P(A) = \int_A f(x)dx$

□ The cumulative distribution function (cdf) is defined as $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$.

□ The expectation is defined as $\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx$

□ The variance is defined as $\sigma^2 = Var(X) = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$

□ The moment generating function is defined as $M(t) = \phi(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$, $-h < t < h$ for some $h > 0$.

□ The (100α)th percentile is x_α such that $F(x_\alpha) = \int_{-\infty}^{x_\alpha} f(x)dx = \alpha$.

♢ *Example:* Let X be the distance in feet between bad records on a used tape with the p.d.f.

$$f(x) = \frac{1}{40}e^{-x/40}, \quad 0 \leq x < \infty$$

Then the probability that no bad records appear within the first 40 feet is

$$P(X > 40) = \int_{40}^{\infty} f(x)dx = e^{-1} = 0.368$$

Uniform, Exponential, Gamma, Chi-Square, Normal, Beta, Student-t, and F Distributions

Uniform $U(a,b)$ $f(x) = \frac{1}{b-a}$, $a \le x \le b$

Exponential $f(x) = \frac{1}{\theta} e^{-x/\theta}$, $0 < x < \infty$

Gamma $f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$, $0 < x < \infty$

$\chi^2(r)$ **Chi-Square** $f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{(r/2)-1} e^{-x/2}$, $0 < x < \infty$

$N(\mu, \sigma^2)$ **Normal** $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$, $-\infty < x < \infty$

Beta Distribution $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0 < x < 1$

Let $Z \sim N(0,1)$, $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$, define $T = \frac{Z}{\sqrt{\chi^2(n)/n}}$ and $F = \frac{\chi^2(n)/n}{\chi^2(m)/m}$, then

Student-t Distribution $f_T(t) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)[1+(t^2/n)]^{(n+1)/2}}$, $\infty < t < \infty$

F Distribution $f_F(w) = \frac{\Gamma((n+m)/2)(n/m)^{n/2} w^{(n/2)-1}}{\Gamma(n/2)\Gamma(m/2)[1+nw/m]^{(n+m)/2}}$, $0 < w < \infty$

Gamma, Exponential, χ^2 Distributions

Consider an (approximate) *Posisson distribution* with mean (arrival rate) λ, let the r.v. X be the waiting time until the αth arrival occurs. Then the cumulative distribution of X can be expressed as

$$\begin{aligned} F(x) &= P(X \leq x) = 1 - P(X > x) \\ &= 1 - P(fewer\ than\ \alpha\ arrivals\ in\ (0, x]) \\ &= 1 - \textstyle\sum_{k=0}^{\alpha-1}[e^{-\lambda x}(\lambda x)^k/(k!)] \end{aligned} \tag{1}$$

$$f(x) = F'(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, 0 < x < \infty, \quad \alpha > 0 \tag{2}$$

Let $\theta = 1/\lambda$, we have the p.d.f. of Gamma Distribution

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, 0 < x < \infty \tag{3}$$

For Gamma distribution, if $\alpha = 1$, we have the p.d.f. of Exponential Distribution

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 < x < \infty \tag{4}$$

For Gamma distribution, if $\theta = 2$ and $\alpha = r/2$, where r is a positive integer, then we have the p.d.f. of $\chi^2(r)$ distribution with r degrees of freedom.

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{(r/2)-1} e^{-x/2}, \quad 0 < x < \infty \tag{5}$$

For Gamma Distribution,

$M(t) = 1/(1 - \theta t)^\alpha$, $\mu = \alpha\theta$, $\sigma^2 = \alpha\theta^2$

For Exponential Distribution,

$M(t) = 1/(1 - \theta t)$, $\mu = \theta$, $\sigma^2 = \theta^2$

For $\chi^2(r)$ Distribution,

$M(t) = 1/(1 - 2t)^{r/2}$, $\mu = r$, $\sigma^2 = 2r$

Normal (Gaussian) Distributions

A normal distribution of r.v. $X \sim N(\mu, \sigma^2)$ has the p.d.f.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \tag{6}$$

□ $M(t) = \phi(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

□ $(X - \mu)/\sigma \sim N(0, 1)$

□ $Z \sim N(0, 1) \;\Rightarrow\; Y = Z^2 \sim \chi^2(1)$

When $\mu = 0$, $\sigma = 1$, $X \sim N(0, 1)$ is said to have the *standard normal distribution*. The cumulative distribution is denoted as

♢ $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad -\infty < z < \infty$

Define $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ for $x > 0$ and let $\gamma = \int_0^\infty e^{-x^2} dx$, show that

(a) $\gamma = \int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$

(b) Show that $\Gamma(\frac{1}{2}) = 2\gamma = \sqrt{\pi}$

(c) $\Gamma(x+1) = x\Gamma(x),\; for\; x > 0,\; \Gamma(n) = (n-1)!$ if $n \in N$.

Proof: (a)

$$\begin{aligned} \gamma^2 &= \left(\int_0^\infty e^{-x^2} dx\right)\left(\int_0^\infty e^{-y^2} dy\right) &&= \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dx dy \\ &= \int_0^{\frac{\pi}{2}} \int_0^\infty e^{-r^2} J_{r,\theta}(x, y) dr d\theta &&= \int_0^{\frac{\pi}{2}} \int_0^\infty e^{-r^2} r dr d\theta \\ &= \frac{\pi}{4} \end{aligned} \tag{7}$$

Proof: (b)

$$\begin{aligned} \Gamma(\tfrac{1}{2}) &= \int_0^\infty e^{-t} t^{\frac{1}{2}-1} dt &&= \int_0^\infty e^{-t} t^{-\frac{1}{2}} dt \\ &= \int_0^\infty e^{-x^2} t^{-x} d(x^2) &&= 2\int_0^\infty e^{-x^2} dx \\ &= \sqrt{\pi} \end{aligned} \tag{8}$$

Moment-Generating Function for Exponential Distribution

(1) Exponential distribution: $f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0,$

$$\int_0^\infty f(x)dx = \int_0^\infty \frac{1}{\theta} e^{-x/\theta} dx = -e^{-x/\theta} \mid_{x=0}^\infty = 1 \tag{9}$$

$$\begin{aligned} \phi(t) \quad = \quad E[e^{tX}] \quad &= \quad \textstyle\int_0^\infty e^{tx} f(x)dx \\ &= \quad \textstyle\int_0^\infty [e^{tx} \frac{1}{\theta} e^{-x/\theta}] dx \\ &= \quad \textstyle\frac{-1}{1-\theta t} e^{-[(1/\theta)-t]x} |_{x=0}^\infty \\ &= \quad \textstyle\frac{1}{1-\theta t} \quad for\ t < \frac{1}{\theta} \end{aligned} \tag{10}$$

$$E[X] = \phi'(0) = \frac{\theta}{(1-\theta t)^2} \mid_{t=0} = \theta \tag{11}$$

$$Var(X) = \phi''(0) - [\phi'(0)]^2 = \frac{2\theta^2}{(1-\theta t)^3} \mid_{t=0} - \theta^2 = \theta^2 \tag{12}$$

Moment-Generating Function for $N(\mu, \sigma^2)$ Distribution

(2) Normal distribution: $X \sim N(\mu,\ \sigma^2)$, $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$, $\quad -\infty < x < \infty$,

(2.1) $\gamma = \int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$.

Proof:

$$\begin{aligned}\gamma^2 &= \left(\int_0^\infty e^{-x^2} dx\right)\left(\int_0^\infty e^{-y^2} dy\right) &&= \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dxdy \\ &= \int_0^{\frac{\pi}{2}} \int_0^\infty e^{-r^2} J_{r,\theta}(x,y) dr d\theta &&= \int_0^{\frac{\pi}{2}} \int_0^\infty e^{-r^2} r dr d\theta \\ &= \frac{\pi}{4}\end{aligned} \tag{13}$$

(2.2) Given $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dx$ for $x > 0$, then $\Gamma(x+1) = x\Gamma(x)$.

(2.3) $\Gamma(n+1) = n!$ for $n \geq 0$, where $0! = 1$, and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

(2.4) $X \sim N(\mu,\ \sigma^2)$, then $\phi(t) = E[e^{tX}] = e^{\mu t + \frac{\sigma^2 t^2}{2}}$.

Proof:

$$\begin{aligned}\phi(t) &= \int_{-\infty}^\infty [e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}] dx \\ &= \int_{-\infty}^\infty \left\{ \frac{1}{\sqrt{2\pi}\sigma} e^{[-(x-(\mu+\sigma^2 t))^2 + (2\sigma^2 \mu t + \sigma^4 t^2)]/2\sigma^2} \right\} dx \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}} \int_{-\infty}^\infty [\frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}] dy \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}}\end{aligned} \tag{14}$$

Moment-Generating Function for $\chi^2(r)$ Distribution

(3) $Z \sim N(0,1) \;\Rightarrow\; Y = Z^2 \sim \chi^2(1)$.

Proof:

$$F(y) = P[Y < y] = P[-\sqrt{y} < Z < \sqrt{y}] = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \tag{15}$$

$$f(y) = F'(y) = \frac{1}{\Gamma(\frac{1}{2})2^{1/2}} y^{-1/2} e^{-y/2}, \quad 0 < y < \infty \tag{16}$$

(3.1) $Y \sim \chi^2(1)$, then $\phi(t) = \frac{1}{\sqrt{1-2t}}$.

Proof:

$$\begin{aligned} \phi(t) &= \int_0^\infty [e^{tx} \tfrac{1}{\Gamma(\frac{1}{2})2^{1/2}} x^{-1/2} e^{-x/2}] dx \\ &= \tfrac{1}{\Gamma(\frac{1}{2})2^{1/2}} \int_0^\infty [e^{-(\frac{1}{2}-t)x} x^{-1/2}] dx \\ &= \tfrac{1}{\sqrt{1-2t}} \tfrac{1}{\Gamma(\frac{1}{2})} \int_0^\infty [e^{-y} y^{-1/2}] dy \\ &= \tfrac{1}{\sqrt{1-2t}} \end{aligned} \tag{17}$$

(3.2) Let $Y_j \sim \chi^2(1),\ 1 \le j \le r$, be independent χ^2 distribution with 1 degree of freedom. Define $Y = \sum_{j=1}^r Y_j$, then $Y \sim \chi^2(r)$ has χ^2 distribution with r degrees of freedom. The p.d.f. and the moment-generating function are given below.

$$f(y) = \frac{1}{\Gamma(r/2)2^{r/2}} y^{(r/2)-1} e^{-y/2}, \quad 0 < y < \infty \tag{18}$$

$$\phi(t) = 1/(1-2t)^{r/2}, \quad E[Y] = \phi'(0) = r, \quad Var(Y) = \phi''(0) - [\phi'(0)]^2 = 2r \tag{19}$$

Proof:

$$\begin{aligned} \phi_Y(t) &= E[exp(tY)] &&= E[exp(\textstyle\sum_{j=1}^r tY_j)] \\ &= \Pi_{j=1}^r E[exp(tY_j)] &&= \Pi_{j=1}^r \phi_{Y_j}(t) \\ &= \Pi_{j=1}^r \tfrac{1}{\sqrt{1-2t}} &&= \tfrac{1}{(1-2t)^{r/2}} \end{aligned} \tag{20}$$

Moment-Generating Functions for Gamma Distributions

(4) *Gamma*(α, θ) distribution: $f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}, \quad 0 < x < \infty$

(4.1) For Gamma distribution, if $\alpha = 1$, we have the p.d.f. of Exponential Distribution

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 < x < \infty \tag{21}$$

(4.2) For Gamma distribution, if $\theta = 2$ and $\alpha = r/2$, where r is a positive integer, then we have the p.d.f. of χ^2 distribution with r degrees of freedom, denoted as $X \sim \chi^2(r)$.

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{(r/2)-1} e^{-x/2}, \quad 0 < x < \infty \tag{22}$$

(4.3) $\phi(t) = 1/(1-\theta t)^\alpha$, $\mu = \alpha\theta$, $\sigma^2 = \alpha\theta^2$

(4.4) For Exponential distribution, $\phi(t) = 1/(1-\theta t)$, $\mu = \theta$, $\sigma^2 = \theta^2$

(4.5) For $\chi^2(r)$ distribution, $\phi(t) = 1/(1-2t)^{r/2}$, $\mu = r$, $\sigma^2 = 2r$

Proof:

$$\begin{aligned}
\phi(t) &= \int_0^\infty \left[e^{tx} \frac{1}{\Gamma(\alpha)\alpha^\theta} x^{\alpha-1} e^{-x/\theta} \right] dx \\
&= \frac{1}{\Gamma(\alpha)\alpha^\theta} \int_0^\infty \left(e^{-(\frac{1}{\theta}-t)x} x^{\alpha-1} \right) dx \\
&= \frac{1}{(1-\theta t)^\alpha} \frac{1}{\Gamma(\alpha)} \int_0^\infty [e^{-y} y^{\alpha-1}] dy \\
&= \frac{1}{(1-\theta t)^\alpha}
\end{aligned} \tag{23}$$

Exercises for $N(\mu, \sigma^2)$, Gamma(α, θ), χ^2 Distributions

1. Define $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ for $x > 0$ and let $\gamma = \int_0^\infty e^{-x^2} dx$, show that

(a) $\gamma = \int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$

(b) Show that $\Gamma(\frac{1}{2}) = 2\gamma = \sqrt{\pi}$

(c) $\Gamma(x+1) = x\Gamma(x), \ for\ x > 0, \ \Gamma(n) = (n-1)!$ if $n \in N$.

2. Let $X \sim N(10, 36)$, write down the pdf for X and compute

(a) P$(X > 5)$.

(b) P$(4 < X < 16)$.

(c) P$(X < 8)$.

(d) P$(X < 20)$.

(e) P$(X > 16)$.

3. Let $Z \sim N(0, 1)$ and $Y = Z^2$, then Y is said to have a χ^2 distribution of 1 degree of freedom, denoted as $\chi^2(1)$.

(a) Show that $f_Y(y) = \frac{1}{\Gamma(1/2)2^{1/2}} y^{-1/2} e^{-y/2}, \ \ 0 < y < \infty$

(b) Show that $\phi(t) = \frac{1}{(1-2t)^{1/2}}, \ \ t < \frac{1}{2}$

(c) If $Y_1, Y_2, \cdots, Y_k \sim \chi^2(1)$ and $Y_1, Y_2, \cdots, Y_k$ are independent, define $W = \sum_{j=1}^k Y_j$, then $W \sim \chi^2(k)$.

(d) Show that $\phi_W(t) = \frac{1}{(1-2t)^{k/2}}$ and $f_W(x) = \frac{1}{\Gamma(k/2)2^{k/2}} x^{(k/2)-1} e^{-x/2}, \ \ 0 < x < \infty$

(e) $Y \sim \chi^2(6)$ is a random variable with 6 degrees of freedom, write down the pdf for Y and compute P$(Y \leq 6)$ and P$(3 \leq Y \leq 9)$.

4. If W is an exponential distribution with mean 6, write down the pdf for W and compute

(a) P$(W < 6)$.

(b) P$(W > 18 \mid W > 12)$.

Matlab Codes for Continuous Distributions

```
% Script file: ch3.m - Continuous Distributions
%
% Exponential Distribution
  subplot(2,2,1)
  X=0.1:0.1:12;
  Ya=exppdf(X,1); Yb=exppdf(X,2); Yc=exppdf(X,4); Yd=exppdf(X,7);
  plot(X,Ya,'r-',X,Yb,'g-',X,Yc,'b-',X,Yd,'m-'); %axis([0,12, 0,0.3])
  legend('Exp(1)','Exp(2)','Exp(4)','Exp(7)')
  title('(4) Exponential(\theta), \theta=1,2,4,7')
%
% Chi-Square Distributions
  subplot(2,2,2)
  X=0.1:0.1:12;
  Y1=chi2pdf(X,1); Y2=chi2pdf(X,2); Y4=chi2pdf(X,4); Y7=chi2pdf(X,7);
  plot(X,Y1,'r-',X,Y2,'g-',X,Y4,'b-',X,Y7,'m-'); %axis([0,12, 0,0.3])
  legend('\chi^2(1)','\chi^2(2)','\chi^2(4)','\chi^2(7)')
  title('(5) \chi^2(r), r=1,2,4,7')
%
% Normal Distributions
  subplot(2,2,3)
  X7=-6:0.2:6;   u=0;  s1=1;  s2=2;  s3=2.5;  s4=3;
  Y7a=normpdf(X7,u,s1); Y7b=normpdf(X7,u,s2); Y7c=normpdf(X7,u,s3);
  Y7d=normpdf(X7,u,s4);
  plot(X7,Y7a,'r-',X7,Y7b,'g-',X7,Y7c,'b-',X7,Y7d,'m-');
  axis([-6,6, 0,0.42])
  legend('N(0,1)','N(0,4)','N(0,6.25)','N(0,9)')
  title('(6) Normal Distribution:  N(u,s^2)')
%
% Gamma Distributions
  subplot(2,2,4)
  X6=0.1:0.1:12; t1=2; t2=3; t3=4; t4=4;
  Y6a=gampdf(X6,3,t1); Y6b=gampdf(X6,3,t2);
  Y6c=gampdf(X6,2,t4); Y6d=gampdf(X6,4,t4);
  plot(X6,Y6a,'r-',X6,Y6b,'g-',X6,Y6c,'b-',X6,Y6d,'m-');
% axis([0,12, 0,0.3])
  legend('\Gamma(3,2)','\Gamma(3,3)','\Gamma(2,4)','\Gamma(4,4)')
  title('(7) \Gamma(\alpha, \lambda)')
%
% Uniform Distributions
% X1=0:0.01:1;  Y1=unifpdf(X1,0,1);  plot(X1,Y1,'r-');
% title((9) 'Uniform Distribution U(0,1)')
```

Some Continuous Distributions