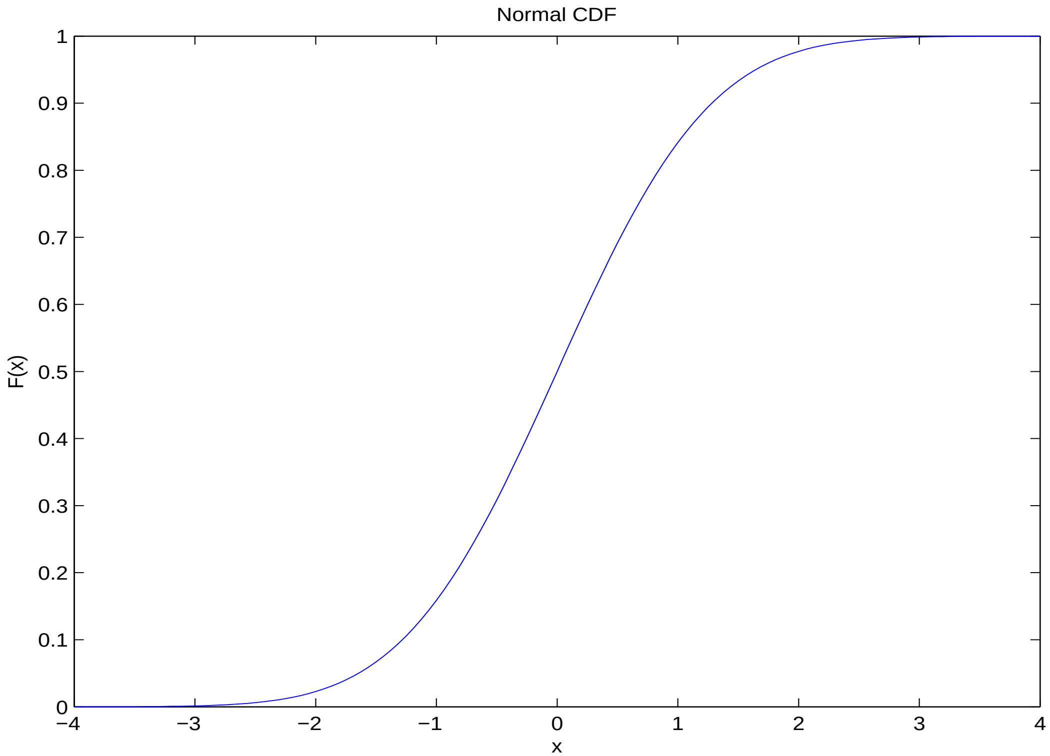
CONTINUOUS RV GENERATION

Inversion Method: for an RV X with pdf $f(x)$, for $x \in [a,b]$.

- Basic method: if $F(x) = \int_a^x f(t)dt$, generate $X = F^{-1}(U)$, with $U \sim Uni(0,1)$.
- Analysis uses $P\{X \leq x\} = F(x) \in [0,1]$;
  F is monotone increasing, so $F^{-1}(u)$ defined and therefore

$$P\{X \leq x\} = P\{F^{-1}(U) \leq x\} = P\{F(F^{-1}(U)) \leq F(x)\} = P\{U \leq F(x)\}.$$

Example: Normal cdf $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$.



- Primary difficulty: given u, how to solve $F(x) = u$ for x.

CONTINUOUS RVs CONTINUED

- Some continous cdfs with easy inversions:
  - *uniform* for $[a, b]$, pdf $f(x) = \frac{1}{b-a}$.

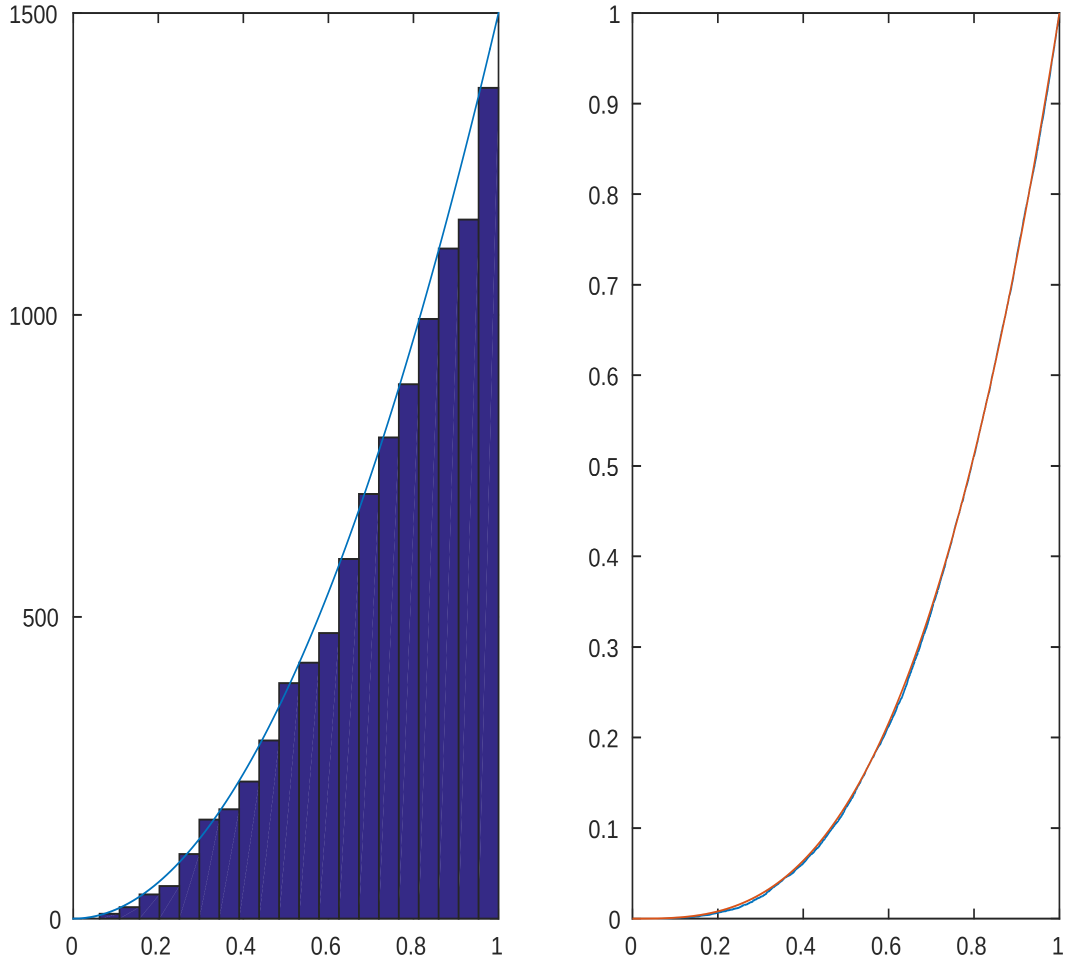
    use $X = a + (b - a)U$;
  - *power* function for $[0, 1]$, pdf $f(x) = nx^{n-1}$.

    use $X = U^{1/n}$;

CONTINUOUS RVs CONTINUED

Matlab Test with power pdf with $n = 3$.

```
K = 10000; X = rand(1,K).^(1/3);
subplot(1,2,1); hist(X,20)
holdon, x = [0:100]/100; plot(x,500*3*x.^2);
subplot(1,2,2); plot( sort(X),[1:K]/K, x,x.^3 )
```



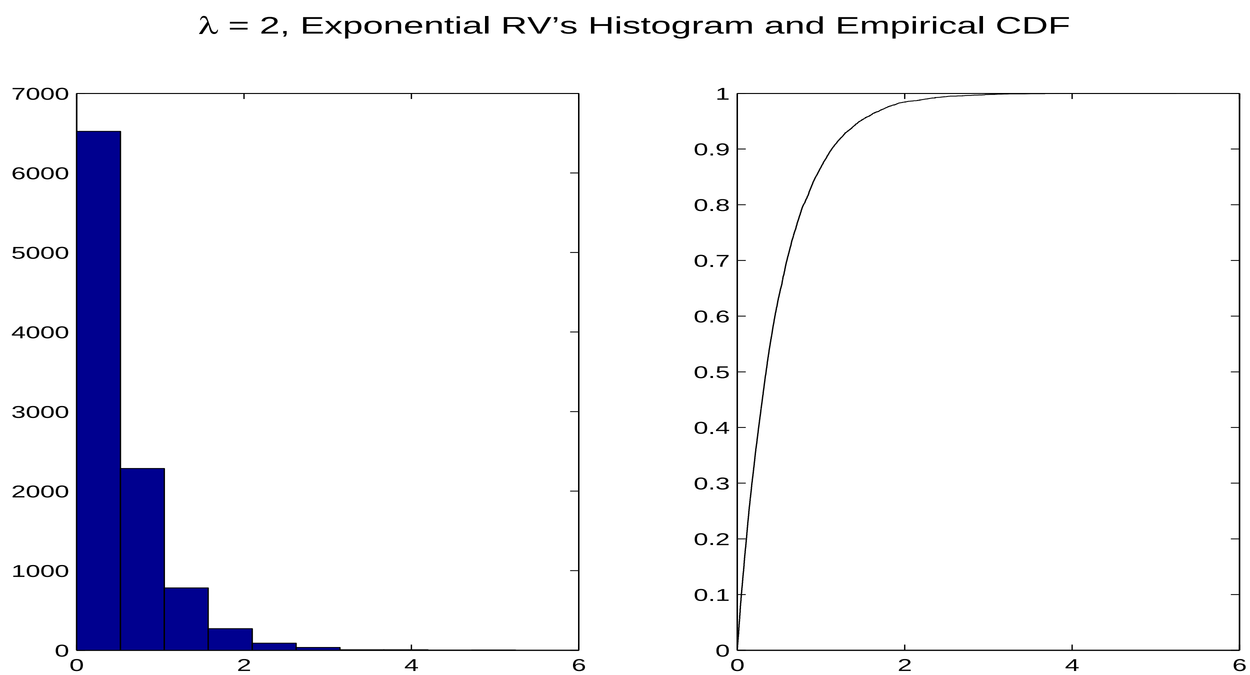
CONTINUOUS RVs CONTINUED

– *exponential* for $[0, \infty)$, pdf $f(x) = \lambda e^{-\lambda x}$

use $X = -ln(1-U)/\lambda$, or $X = -ln(U)/\lambda$.

Matlab example: for $\lambda = 2$, compute 10000 X's, plot empirical cdf.

```
K = 10000; X = -log(rand(1,K))/2; disp(mean(X)) ... 0.49916
subplot(1,2,1); hist(X), subplot(1,2,2); plot( sort(X),[1:K]/K )
```



CONTINUOUS RVs CONTINUED

- Some cdfs with no easy formula inversions:
  - *Beta(a,b)* for $[0,1]$, $F(x) = K_{a,b}\int_0^x t^{a-1}(1-t)^{b-1}dt$;
  - *Gamma(a,b)* for $[0,\infty)$, $F(x) = K_a\int_0^x b(bt)^{a-1}e^{-bt}dt$;
  - *Normal* for $(-\infty,\infty)$, $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-t^2/2}dt$.
- Numerical Inversion; e.g. Newton's method: solve $F(X) - U = 0$; starting with X_0, and iterating $X_{i+1} = X_i - (F(X_i) - U)/f(X_i)$, until convergence.

  Example: Normal RV's

  Matlab

```
U = rand, X = .5; % Newton Method starting guess
for i = 1 : 4, X = X-(normcdf(X)-U)/normpdf(X); disp([i X]) end
 disp(norminv(U)) % Check
U = 0.75672555764055
        1            0.685372121038858
        2            0.695770886529439
        3            0.695808308958099
        4            0.695808309445309
ans = 0.695808309445309
```

  Too slow for most applications; and so is $\Phi^{-1}(U)$.

CONTINUOUS RVs CONTINUED

- Generation via Transformation to easily inverted cdfs:
  - Poisson RV N, with pmf $p_j = e^{-\lambda}\lambda^j/j!, j = 0, 1, \ldots:$

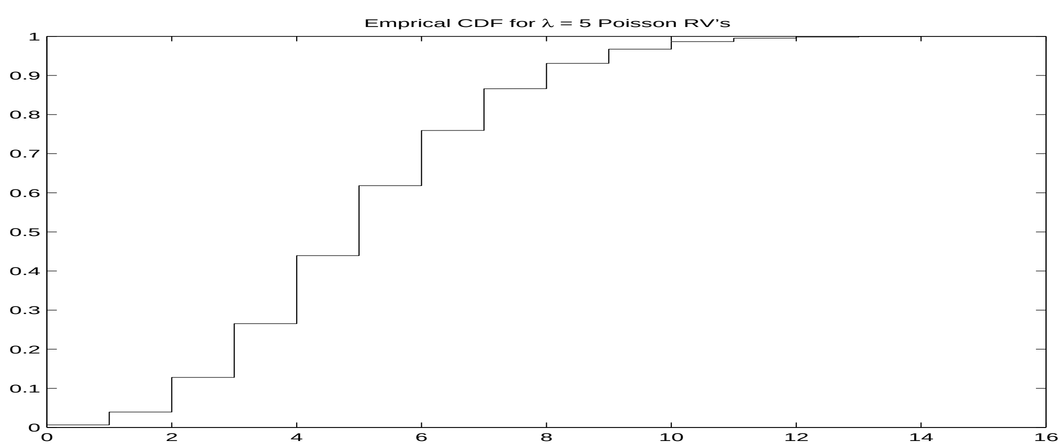
    $N = Max\{n : \sum_{i=1}^{n} X_i \le 1\}$, X_is exponential, mean $1/\lambda$.

    So $N = Max\{n : -\sum_{i=1}^{n} \ln(U_i)/\lambda \le 1\}$

    $$= Max\{n : \ln(\prod_{i=1}^{n} U_i) \ge -\lambda\} = Max\{n : \prod_{i=1}^{n} U_i \ge e^{-\lambda}\}.$$

    Algorithm: generate U_i's until $\prod_{i=1}^{n} U_i < e^{-\lambda}$, set $N = n - 1$. Matlab

```
K = 10000; lam = 5; elam = exp(-lam);
for i = 1 : K, n = 1; p = rand;
  while p > elam, p = p*rand; n = n + 1; end, X(i) = n - 1;
end, disp([mean(X) lam]), plot(sort(X),[1:K]/K)... 4.9937   5
```



CONTINUOUS RVs CONTINUED

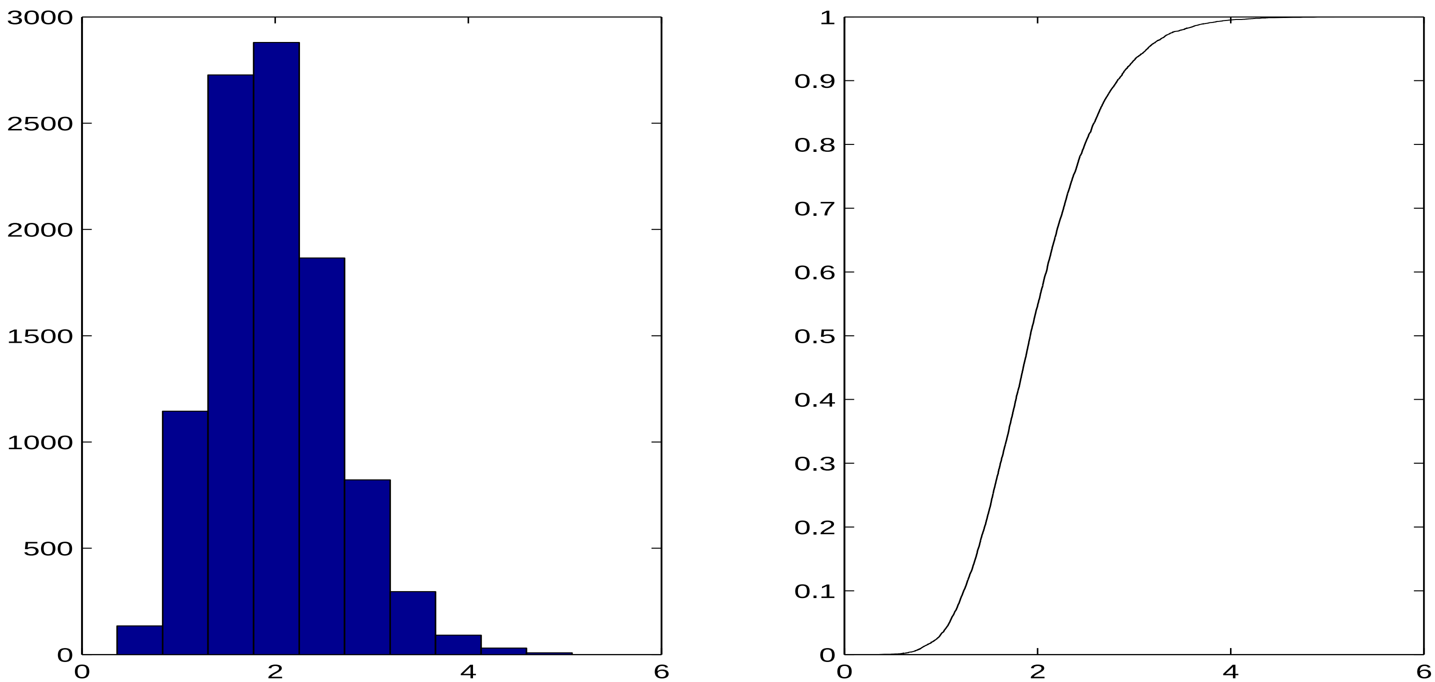
– Gamma(n,b) RV X, with pdf $f(x) = K(bx)^{n-1}e^{-bx}$, $x > 0$, with integer n:

$X = \sum_{i=1}^{n} X_i$, if X_i's are exponential with mean $1/b$.

So $X = -\sum_{i=1}^{n} \ln(U_i)/b = -\ln\left(\prod_{i=1}^{n} U_i\right)/b$;

```
% Matlab
K = 10000; n = 10; b = 5; X = -log( prod( rand(n,K) ) )/b;
subplot(1,2,1); hist(X), subplot(1,2,2); plot( sort(X),[1:K]/K )
disp([mean(X) n/b])   1.994            2
```

n = 10, b = 5, Gamma RV's Histogram and Empirical CDF



CONTINUOUS RVs CONTINUED

Acceptance-Rejection Method (AR):

- Assume an efficient method is available for generating Y RVs, with Y pdf $g(y)$, and $f(y) \leq cg(y)$, $\forall y$, and for some c.
- AR Algorithm:
  1) generate Y, with pdf g;
  2) generate $U \sim uni(0,1)$;
  3) if $U < \frac{f(Y)}{cg(Y)}$, set $X = Y$, else go to step 1.
- Analysis uses

$$P\{accepted\} = \int P\{accepted|X = y)g(y)dy$$
$$= \int \frac{f(y)}{cg(y)} g(y)dy = 1/c, \text{ so}$$

$$P\{X = x|x \ accepted\} = \frac{P\{X{=}x \ and \ U{<}\frac{f(Y)}{cg(Y)}\}}{P\{accepted\}}$$
$$= \frac{f(x)}{cg(x)} g(x)c = f(x).$$

Note: usually c is chosen as $c = max_y(f(y)/g(y))$, so average # of steps is c.

CONTINUOUS RVs CONTINUED

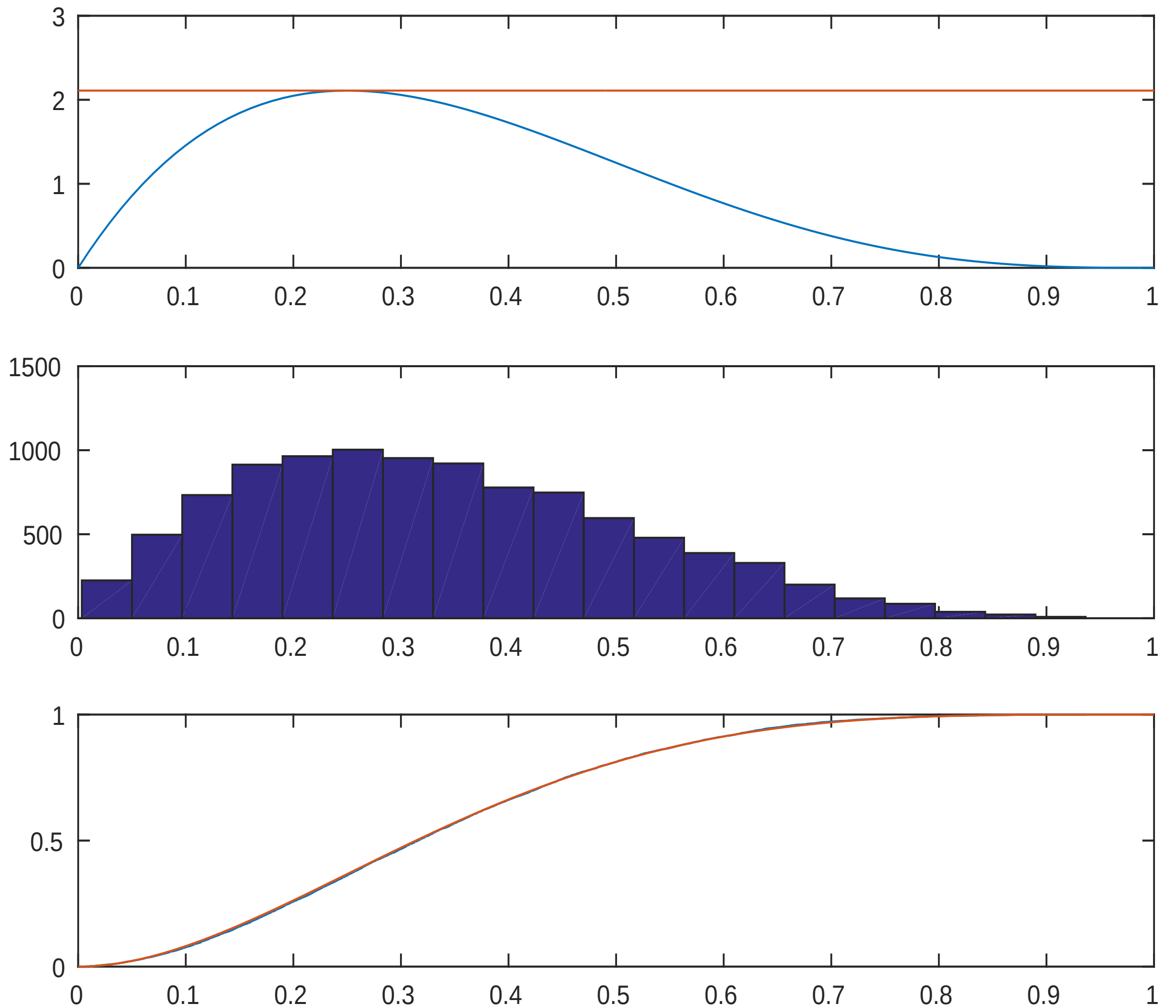
- AR Beta Example: Beta(2,4) pdf $f(x) = 20x(1-x)^3$, $0 < x < 1$, with $\mu = 1/3$; notice $F(x) = 10x^2 - 20x^3 + 15x^4 - 4x^5$, so no direct inversion; try $g(x) = 1,\ c = max_x(f(x)) = f(1/4) = 135/64 \approx 2.11$.

```
K = 10000; c = 135/64; f = @(x)20*x.*(1-x).^3;
for i = 1 : K, k = 1; Y = rand; % from g(x) = 1
  while rand > f(Y)/c, k = k + 1; Y = rand; end
  X(i) = Y; C(i) = k;
end, disp([ mean(X) mean(C) ])
    0.33343       2.0989
```

CONTINUOUS RVs CONTINUED

Beta Graphs

```
subplot(3,1,1); x = [0:.01:1]; plot(x,f(x),x,c*ones(1,101));
subplot(3,1,2), hist(X,20),   subplot(3,1,3)
plot( sort(X),[1:K]/K,  x,x.^2.*(10-20*x+15*x.*x-4*x.^3) )
```



CONTINUOUS RVs CONTINUED

- AR Gamma Example:
  Gamma(3/2,1) pdf $f(x) = \frac{2}{\sqrt{\pi}} x^{1/2} e^{-x}$, $x > 0, \mu = 3/2$;
  try an exponential $g(x)$ with the same mean $g(x) = \frac{2}{3} e^{-2x/3}$, then
  $c = max_x(f(x)/g(x)) = f(3/2)/g(3/2) = 3\sqrt{\frac{3}{2e\pi}} \approx 1.26$.

Matlab

```
K = 10000; c = 3*sqrt(3/(2*pi*exp(1)));
f = @(x)2*sqrt(x/pi).*exp(-x); g = @(x)2*exp(-2*x/3)/3;
for i = 1:K, k = 1; Y = -3*log(rand)/2;
  while rand > f(Y)/(g(Y)*c)
    k = k + 1; Y = -3*log(rand)/2;
  end, X(i) = Y; C(i) = k;
end, disp([ mean(X) mean(C) ])
       1.5003       1.2448
```

CONTINUOUS RVs CONTINUED

Gamma Graphs

```
subplot(3,1,1); x = [0:.05:12]; plot(x,f(x),x,c*g(x));
subplot(3,1,2), hist(X,50)
subplot(3,1,3), plot( sort(X),[1:K]/K,  x,gammainc(x,3/2) )
```