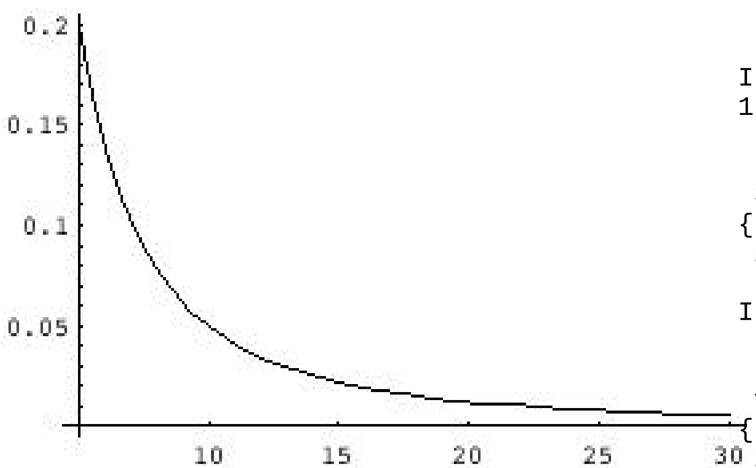
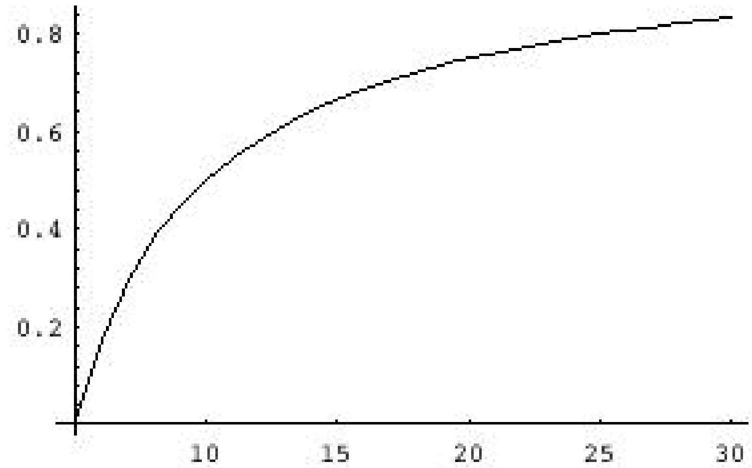
Q4.7

```
In[21]:= f = b y^(-2)

b
--
 2
y

In[27]:=area=Integrate[f,{y,b,Infinity}]

1

In[28]:=F=Simplify[ Integrate[b t^(-
2),{t,b,y}] ]

    b
1 - -
    y

In[29]:= probYGTc =
Simplify[ Integrate[b t^(-
2),{t,b+c,Infinity}] ]

  b
-----
b + c

In[35]:=Plot[(f/.b->5),{y,5,30}]
```

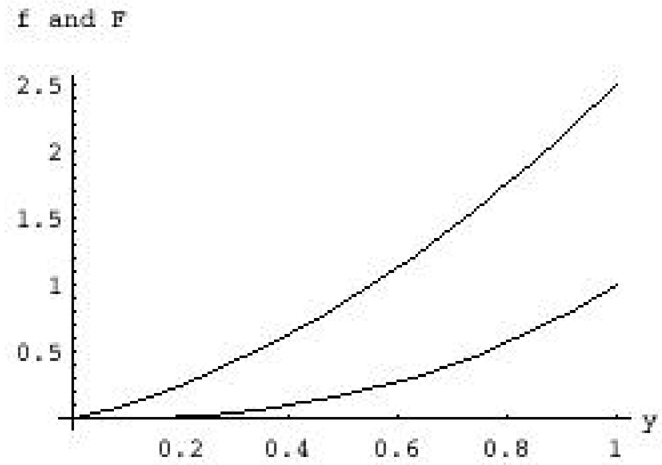



```
In[36]:=Plot[(F/.b->5),{y,5,30}]
```



== **Q4.9** ===

```
In[1]:=f = c y^2 + y

         2
y + c y

In[2]:=i = Integrate[f,y]

 2      3
y    c y
-- + ----
      3

In[6]:=c=c/.Solve[(i/.y->1 - i/.y->0) ==
1,c]


 3
{-}
 2

In[9]:=F = Integrate[(f/.y->t),{t,0,y}]

  2    3
 y    y
{-- + --}
 2    2

In[14]:=Plot[{f,F},{y,0,1},
AxesLabel->{"y","f and F"}]
```




```
F/.y->{-1,0,1,0.5}


{{0, 0, 1, 0.1875}}

In[18]:=
(1-F/.y->0.5)/(1-F/.y->0.25)

{0.845528}

In[20]:=104./123

0.845528
```

==4.12==

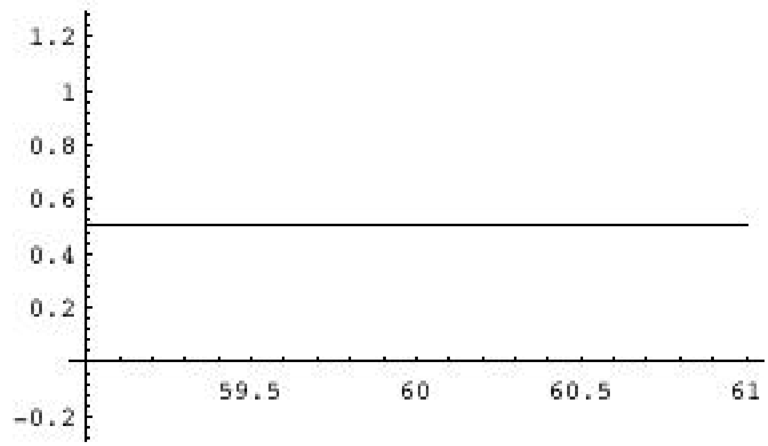
```
In[38]:=
f = (1/Integrate[(2-y),{y,0,2}]) (2-y)

2 - y
-----
  2

In[39]:=
EY=Integrate[y f,{y,0,2}];

VY = Integrate[(y^2) f,{y,0,2}] - (EY)^2;

{EY,VY}//N


{0.666667, 0.222222}
```

==4.13==

```
In[3]:=f = (3/2) y^2 + y;
In[26]:=

EY=Integrate[y f,{y,0,1}];

VY = Integrate[(y^2) f,{y,0,1}] - (EY)^2;

{EY,VY}//N

{0.708333, 0.0482639}
```

==4.19==

```
In[44]:=
f = 0.5; Plot[f,{y,59,61}]
```



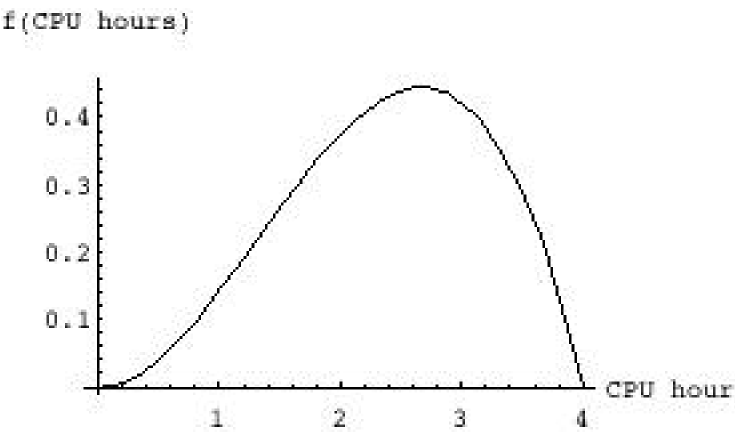

```
In[45]:=

EY=Integrate[y f,{y,59,61}];

VY = Integrate[(y^2) f,{y,59,61}]-(EY)^2;

{EY,VY}//N


{60., 0.333333}
```

NOTE (JH) Var(Uniform) = $\frac{\text{range}^2}{12}$

here range = 2, so V = $\frac{2^2}{12}$ =1/3

==4.22==

```
In[56]:=
f = (3/64) (y^2) (4-y) ; Plot[f,{y,0,4},
AxesLabel->{"CPU hours","f(CPU
hours)"}]
```




```
In[50]:=
EY=Integrate[y f,{y,0,4}];

VY = Integrate[(y^2) f,{y,0,4}] - (EY)^2;

{EY,VY}//N

{2.4, 0.64}
```

E(CPU cost) = \$200 x E(CPU hours)

V(CPU cost) = $\$200^2$ x V(CPU hours)

Prob[cost > \$600) = Prob(CPU hours > 3)

```
In[60]:= ProbGT3=Integrate[f,{y,3.,4.}]

0.261719
```

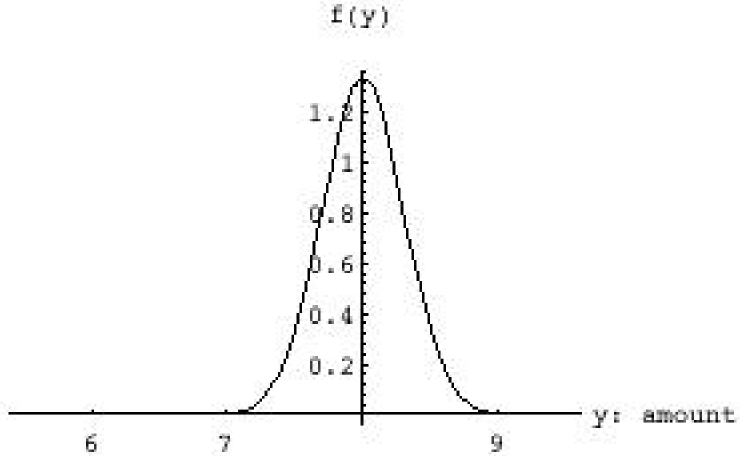
==4.47== Corrected May 17

Y ~ N(μ=??, σ=0.3) for

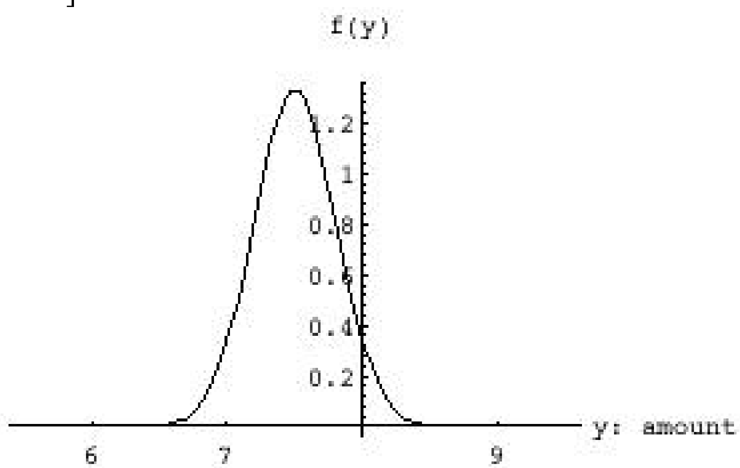
P(Y > 8) = 0.01

A LITTLE TRIAL AND ERROR...

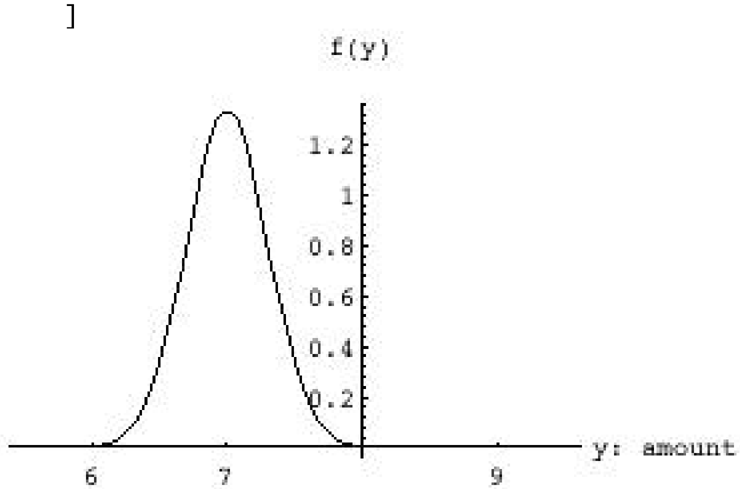
```
f = (1/(sd Sqrt[2 Pi])) *
     Exp[-0.5 ((y - mu)/sd)^2]/.sd->0.3;
```




```
Plot[f/.mu->7.5, {y,5.5,9.5},
AxesLabel->{"y: amount","f(y)"},
PlotRange->All,
AxesOrigin->{8,0}
    ]
```




```
Plot[f/.mu->7.0, {y,5.5,9.5},
AxesLabel->{"y: amount","f(y)"},
PlotRange->All,
AxesOrigin->{8,0}
    ]
```



FROM TABLE INSIDE FRONT COVER OF WMS5..

1% of probability mass is above

z = 2.33 (approx)

Ie if set μ so 8 oz is 2.33 SD's above it, then wil have overflow in 1% of cases

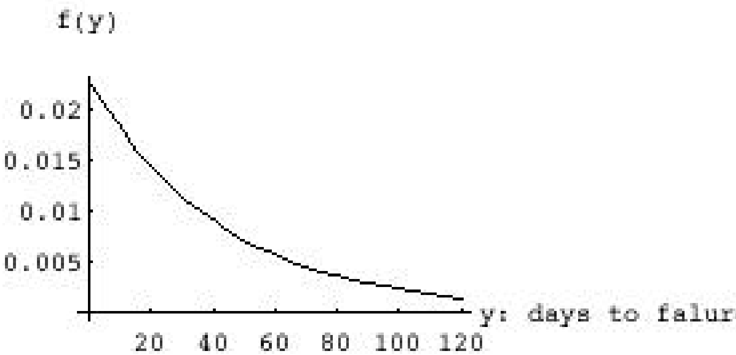
2.33 SDs is 2.33 x 0.3 = 0.699 oz above μ

so.. μ = 8 - 0.699 oz = 7.3 oz.

==4.55==

The "clock" starts again aon July 1. The distribution of times to failure is (neg) exponential with mean m = b = 44 days.

```
f = (1/44) Exp[ - (1/44) y];

Plot[f,{y,0,120},
AxesLabel->{"y: days to falure","f(y)"}]
```




```
F = (Integrate[f,{y,0,t}])/.t->y

     -y/44
1 - E

F/.y->30.0

0.494303

F/.y->31.0
```

0.505667 (Book must have used 31 days)

==4.64==

We are given:

• Gamma Distribuion , which is characterized by parameters α and β (α determines curvature, and α scale), along with its mean 4 sec and variance 8 sec^2

• the values of α and β determine the mean (expected value), the variance, and all higher moments (e.g. avearge cubed deviation from the mean, ...)

Know (or would be given in exam)

$E(Y) = \alpha\beta \quad Var(Y) = \alpha\beta^2$

So can match the 1st 2 "moments" (mean αβ and Var($\alpha\beta^2$) with their numerical counterparts, and solve for α and β (as I did in the breast cancer data analysis)

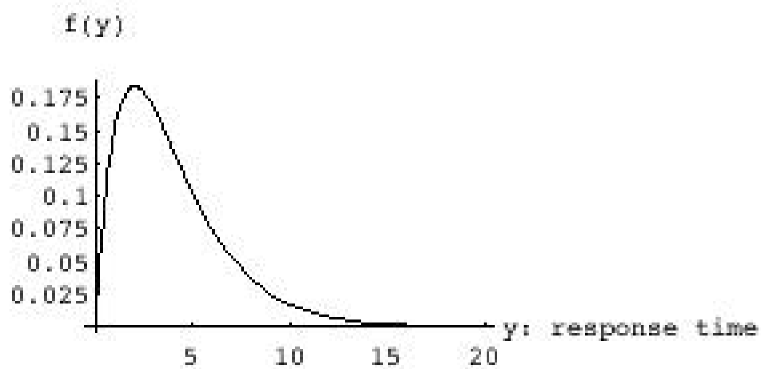
(1)	(2)	(3)	(4)
mean	variance	(2)/(1)	(1)/(3)
$\alpha\beta$	$\alpha\beta^2$	β	α
4	8	2	2

So, from definition 4.8, we have the robability density function

$$(y) = \frac{y^{2-1}\ e^{-y/2}}{2^2\ \Gamma(2)} = \frac{y^1\ e^{-y/2}}{4} \text{ on } [0,\text{Infinity}]$$

heck:

```
f = y Exp[-y/2]/4;

Plot[f,{y,0,20},
AxesLabel->{"y: response time","f(y)"}]
```



Area under curve seems to be unity (if we approximate by a rectangle with (average) height of 0.1 and base of 10!!

Check area (total probability) more closely...

```
area = Integrate[f,{y,0,30}]//N
```

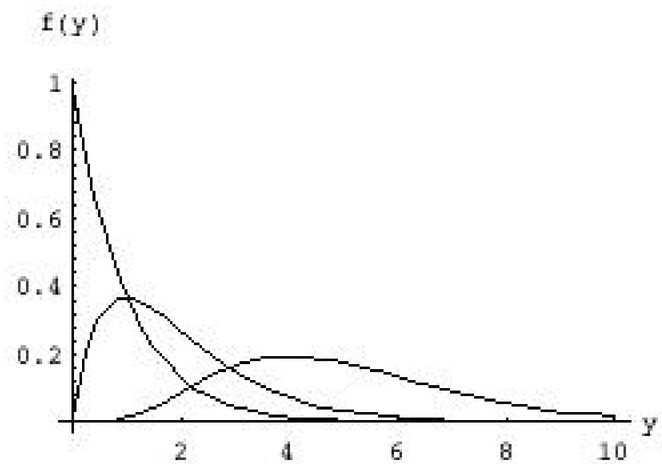
```
0.999995

area = Integrate[f,{y,0,100}]//N

1.
```

NOTE: this distribution is the same as the one for α=2 shown on Fig 4.15 p 159... the only difference is the scale factor β ..

The text must have used a long mean (β) per <u>sub</u>unit" (b), even with the α = 2 subunits.

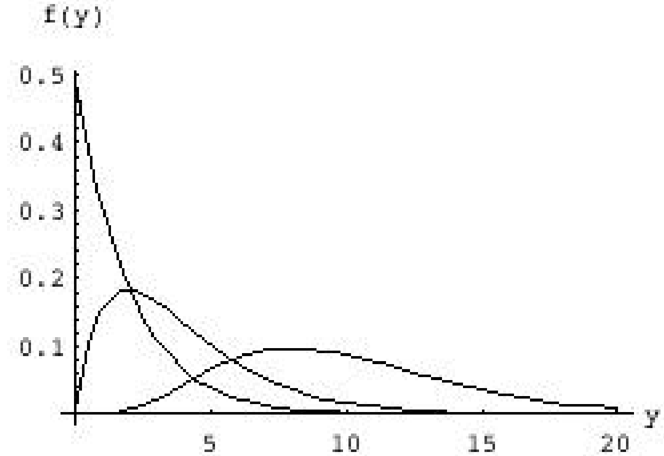
If you took

```
f = (1/(b^a Gamma[a])) y^(a-1) Exp[-
y/b]/.b->1;

Plot[{(f/.a->1),
      (f/.a->2),
      (f/.a->5)},{y,0,10},
PlotRange->All,
AxesLabel->{"y","f(y)"}]
```



if you took

```
f = (1/(b^a Gamma[a])) y^(a-1) Exp[-
y/b]/.b->2;

Plot[{(f/.a->1),
      (f/.a->2),
      (f/.a->5)},{y,0,20},
PlotRange->All,
AxesLabel->{"y","f(y)"}]
```



==4.66==

We are given:

• Gamma Distribuion , which is characterized by parameters α and β (α determines curvature, and α scale), along with

• the values of α and β -- which determine the mean (expected value), the variance, and all higher moments (e.g. avearge cubed deviation from the mean, ...)

$E(Y) = \alpha\beta$ $Var(Y) = \alpha\beta^2$

So can derive the 1st 2 "moments" (mean $\alpha\beta$ and Var($\alpha\beta^2$)

α	β	mean $\alpha\beta$	variance $\alpha\beta^2$	SD ($Var^{1/2}$)
1000	20	\$20,000	(\$^{2}) 400,000	\$630

First off... it is difficult to imagine such a "<u>tight</u>" distribution of income.. a Sd of ONLY \$630 around a mean of \$20,000. No matter what the distribution, this is tight. Even allowing for the "worst" (most contrary, not any "off the shelf, no-name brand") distribution possible, Tchebysheff's theorem says that

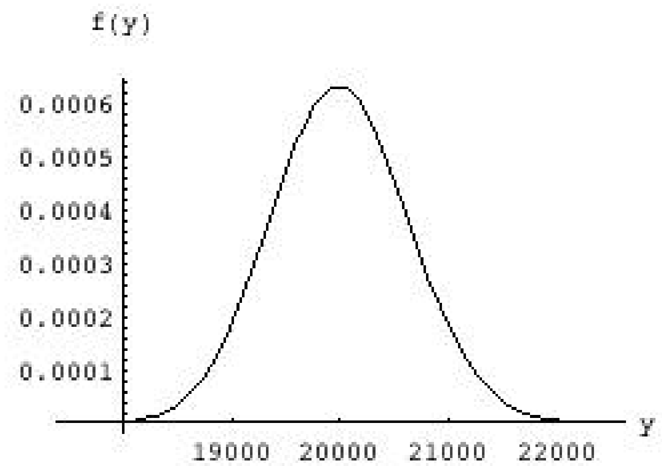
at most $1 - 1/k^2$ of ANY distribution is more than k SD's from the mean ...

So here

at most 1 - 1/4 is > 2(630) from 20,000
at most 1 - 1/9 is > 3(630) from 20,000
at most 1 -1/16 is > 4(630) from 20,000

MOREOVER, as some found when they tried to drwa this distribution in Excel, there are numerical problems trying to compute it.. Mathematica handles it wit aplomb...

```
f = (1/(b^a Gamma[a]))
        y^(a-1) Exp[-y/b]/.b->20;

Plot[(f/.a->1000),{y,17500,22500},
AxesLabel->{"y","f(y)"}]
```



Is it a surprize that it has a Gaussian shape?

NO, if one remembers that a gamma distribution with parameters α and β refers to the distribution of the sum of α independently and identically distributed ("i.i.d.") random variables, each with a negative) exponential distribution with mean β.

The CENTRAL LIMIT THEOREM (vide infra, p 305) says that a sum (or everage) of a lot of i.i.d. random variables will have a close to Gaussian distribution, EVEN if each of these random variables do not themselves have a Gausian distribution.

In our case we are saying that

if Y_i ~ Exp[b], indpenedently of Y_j, then

$\Sigma\ Y_i$ ~ Gaussian (approx) if Σ is over enough rv's.

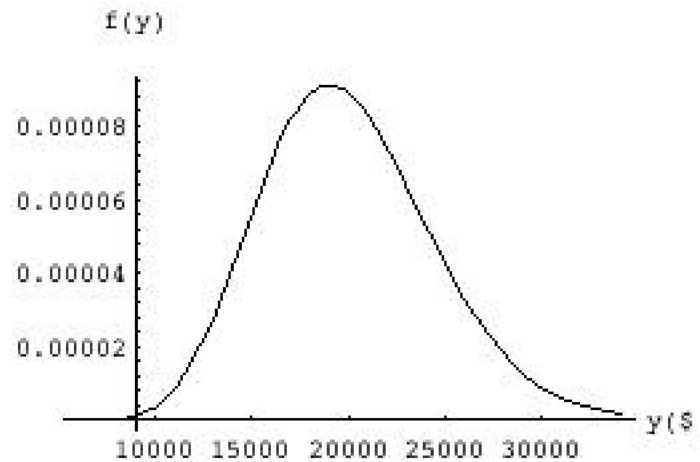
WHAT I THINK THE BOOK MEANT TO ASK...

α	β	mean α β	variance $\alpha\beta^2$	SD ($Var^{1/2}$)
20	1000	$20,000	($^{2})20,000,000	$4,500

i.e. a much more reasonable spread...

Now the distribution looks like...

```
f = (1/(b^a Gamma[a])) y^(a-1) Exp[-
y/b]/.b->1000;

Plot[(f/.a->20),{y,6000,34000},
AxesLabel->{"y","f(y)"}]
```



==4.108==

We are given:

• Poisson Distribution of event <u>count</u> in an interval t, which is characterized by parameter μ = product of mean per time unit (λ) and length of interval (here t units) Note that whereas previously in book, λ was the eman for the interval (<u>whatever</u> length) in question, the book now seems to be switching to using λ as the mean number <u>per time unit</u>.

• T = time to 1st arrival

?? pdf for T

think of it as... (A = arrival)
(sorry I can't make fractional *'s)

```
-------------------------------------time
   A   A        A    A            A
*****
     ****
         *********
                  *****
                       *************
```
etc..

each ****** is a realization of the RV T.

It is difficult to find the pdf of T directly, but, as we often find, it is a lot easier to find the cdf, Prob(T ≤ t), or its complement P(T > t), and to then take the derivative to arrive at the pdf

$f(t) = F'(t)$

Here, having T > t is saying that there is no (zero) arrival in the interval [0,t]. i.e.,

T > t <===> # Arrivals in [0,t] = 0.

So P(T > t)

= PoissonProb(0 events | $\mu = \lambda t$)

$= e^{-\mu} = e^{-\lambda t}$

So...

F(t)

= Prob(T ≤ t)

= 1 - Prob (T > t)

$= 1 - e^{-\lambda t}$

So...

$f(t) = F'(t) = \lambda e^{-\lambda t}$

Now, if the the average number of arrivals is λ per time unit, that means that the average interval betwen arraivals is te reciprocal of this, ie.

$E(T) = 1/ \lambda$,

or

$\lambda = 1/ E(T)$

so

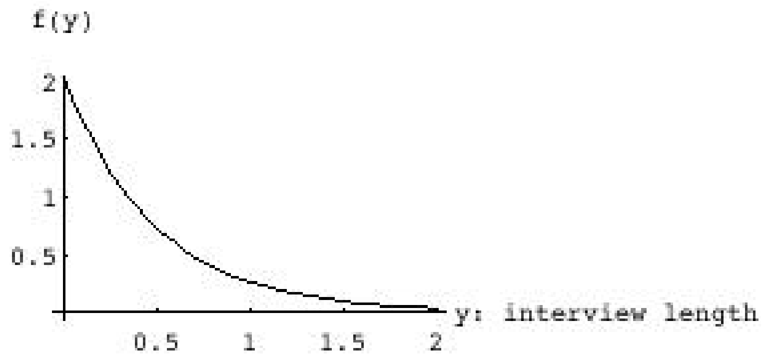
$f(t) = [1/ E(T)]\, e^{- [1/ E(T)]\, t}$

as per the formula for the (neg) exponential pdf function.

== 4.112 == the rude interviewer

Y = length of interview

```
f = (1/0.5) Exp[ - (1/0.5) y ];

Plot[f, {y,0,2},
AxesLabel->{"y: interview length",
            "f(y)"}
    ]
```



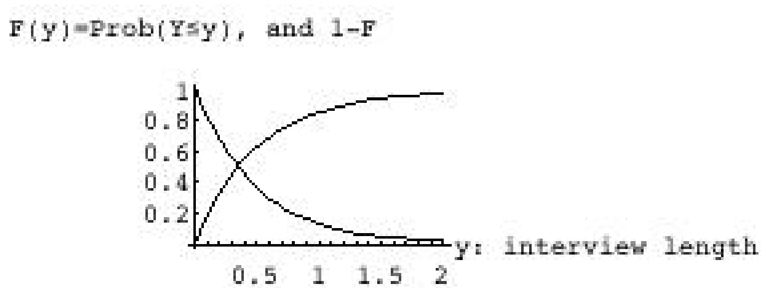
Prob(interview runs past 15 min [1/4 hr])

= Prob(Y > 0.25)

= area under density curve, beyond 0.25

= a sizable probability mass..

```
F = Integrate[f,{y,0,t}]/.t->y

     -2. y
1 - E

Plot[{F,1-F}, {y,0,2},
AxesLabel->{"y: interview length",
            "F(y)=Prob(Y≤y), and 1-F"} ]
```




```
F/.y->0.25

0.39
```

The memoryless ("clueless or ruthless"??) interviewer whose proabability of dismissing the applicant in interval (t,t+Δt), given that interview has already lasted a time t, is the same no matter the value of t!

Compute the conditional "dismissal rates", at 5 minute intervals...

```
Table[{ min,

(F/.y->(min/60)),

1-F/.y->(min/60),


((F/.y->((min+1)/60)) - (F/.y->((min)/60)))
-----------------------------------------
         (1-F/.y->(min/60) )},


{min,5,120,5}]
```

min	F	1-F	Conditional Dismissal Rate $\frac{F(min+1) - F(min)}{1-F(t)}$
5.	0.15	0.85	0.033
10.	0.28	0.72	0.033
15.	0.39	0.61	0.033
20.	0.49	0.51	0.033
25.	0.57	0.43	0.033
30.	0.63	0.37	0.033
35.	0.69	0.31	0.033
40.	0.74	0.26	0.033
45.	0.78	0.22	0.033
50.	0.81	0.19	0.033
55.	0.84	0.16	0.033
60.	0.86	0.14	0.033
65.	0.89	0.11	0.033
70.	0.9	0.097	0.033
75.	0.92	0.082	0.033
80.	0.93	0.069	0.033
85.	0.94	0.059	0.033
90.	0.95	0.05	0.033
95.	0.96	0.042	0.033
100.	0.96	0.036	0.033
110.	0.97	0.03	0.033
110.	0.97	0.026	0.033
110.	0.98	0.022	0.033
120.	0.98	0.018	0.033

WHATEVER the time elapsed in the interview up to now.. there is a 3.3% chance of being dismissed in the next minute ..

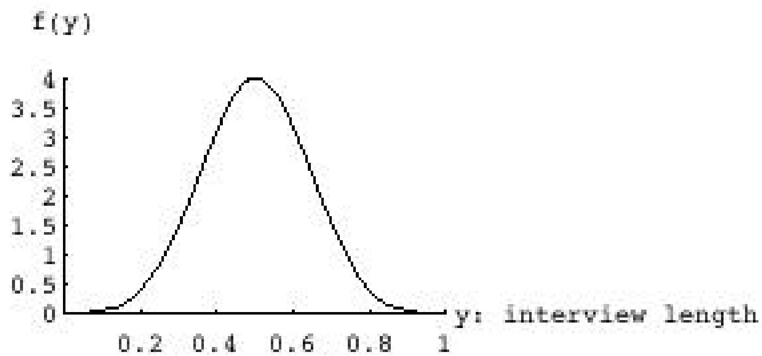
ie a 3.3% chance of being dismissed in 1st min.

Of the ones who "survive" to minute 5 (85% of those who start out), 3.3% of these 85% will be dismissed in the next minute..

Of the ones who "survive" to minute 30 (37% of the 100 who start out), 3.3% of these 37% will be dismissed in the next minute.. etc

What if the durations of interviews had a **Gaussian** distribution with a mean of μ = 1/2 = 0.5 hrs, and a standard deviation of σ = 0.2 hours?

```
f = (1/(sd Sqrt[2 Pi])) *
     Exp[-0.5 (y - mu)^2 /
            (2 sd^2) ]/.{mu->0.5,sd->0.1};

Plot[f, {y,0.0,1.0}],
PlotRange->{{0,1},{0,4}},
AxesLabel->{"y: interview length",
            "f(y)"}
    ]
```



min	F	1-F	Conditional Dismissal Rate $\frac{F(min+1) - F(min)}{1-F(t)}$
5.	0.000015	1.	0.000016
10.	0.00043	1.	0.00034
15.	0.0062	0.99	0.0036
20.	0.048	0.95	0.02
25.	0.2	0.8	0.063
30.	0.5	0.5	0.13
35.	0.8	0.2	0.22
40.	0.95	0.048	0.3
45.	0.99	0.0062	0.38
50.	1.	0.00043	0.46
55.	1.	0.000016	0.52
60.	1.	5.7 10^{-7}	0.29

a **very different picture!**