WMS5 start with the gamma distribution and go on to the exponential as a special case.. In fact, it is easier, and more logical, to arrive at the gamma from the exponential...

The Exponential Distribution is used to describe the variability of..

the time (Y), measured on a continuous scale, until an event (e.g. time to "failure"), where the conditional probability of the event ("failure") in the next time instant, given "survival" (it has not failed) up to this timepoint, is independent of the timepoint

- **The event does not have to be a "failure" .. it could be something good, such as the amount of time until one (finally!) wins the lottery, or gets through to / reach a busy operator or support line, or a busy (or seldom in!) professor**
- **Discrete counterpart is the geometric distribution (§3.5) (n the lottery example, Y is actually discrete!)**
- **Distribution is appropriate if cause of "failure" is "external" and thus independent of age of component**

e.g. Y = longevity of

cups, saucers and plates in cafeteria (main cause of failure is a person, not the item or its age)

some electronic components (e.g. if no moving parts, or if no wear-out)

satellites, the main threat to which is meteors, solar storms, etc..

Y = durations (lengths of intervals) between events which happen with a "memoryless" pattern

memoryless:

P(event in next Δt | event hasn't happened up to t)

= some constant, independent of t

- **Seems to fit well the lengths of time patients with advanced (?terminal) cancer survive after entering studies of "last-ditch" treatments. Why???**

Y ~ Exponential, indexed by a single parameter, β

(?? is index β a meaningful (memorable) parameter ??)

Let's index (characterize) the distribution by its mean μ

Probability density function (pdf).. via F(y)=cdf(y)

As is often the case, it is easier to first derive the cumulative distribution function cdf(y) =F(y) = Prob(Y$\le$y), and is equally often the case, it is often easier to work out the complement, 1 - F(y), of F(y) than F(y) itself.

A mean interval between events of μ implies that, on average, there is 1 event per μ time units.

Key: Divide each time unit into a large number of smaller sub-units or "clicks" (C) say C=100 or C=1000 or C=1000000.

If the mean time to an event is μ (i.e. 1 event per μC clicks], then the average proportion of clicks that generate an event is is the small quantity (probability)

$$\frac{1}{\mu C},$$

and the probability that a click does not generate an event is

$$\left[1-\frac{1}{\mu C}\right].$$

The "memory-less" nature of the process means that, no matter what is "on the clock", the probability of an event in the next "click" is independent of how many clicks have elapsed since we (re)started the clock.

Now the time value y corresponds to yC clicks from the beginning. Thus we can translate the statement "Y > y" into its equivalent

" Y > y " <=======> "no event in yC clicks"

So, ...

Prob(Y > y) = Prob(no event in yC clicks)

$$= \left[1 - \frac{1}{\mu C}\right]^{yC} = \left\{ \left[1 - \frac{1/\mu}{C}\right]^{C} \right\}^{y}$$

(taking limits as C->infinity)

$$= \left\{ e^{-1/\mu} \right\}^{y} = e^{-[1/\mu]\,y}$$

i.e.,

$$F(y) = 1 - \mathrm{Prob}(Y > y) = 1 - e^{-[1/\mu]\,y}$$

so...

$$f(y) = F'(y) = [1/\mu]\, e^{-[1/\mu]\,y}$$

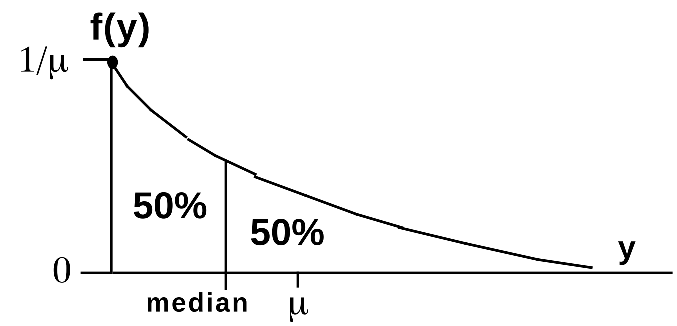
This is the same as the formula in the book, but with our μ replaced by the book's β. Thus, the parameter β in the version in the book is in fact the <u>MEAN</u> of the exponential random variable. [Instead of simply stating what the pdf is, and having to "prove" Theorem 4.10 (namely that E(Y)=β], we have worked out the pdf from first principles -- directly in terms of the mean! This more logical way to doing things also makes it more obvious that

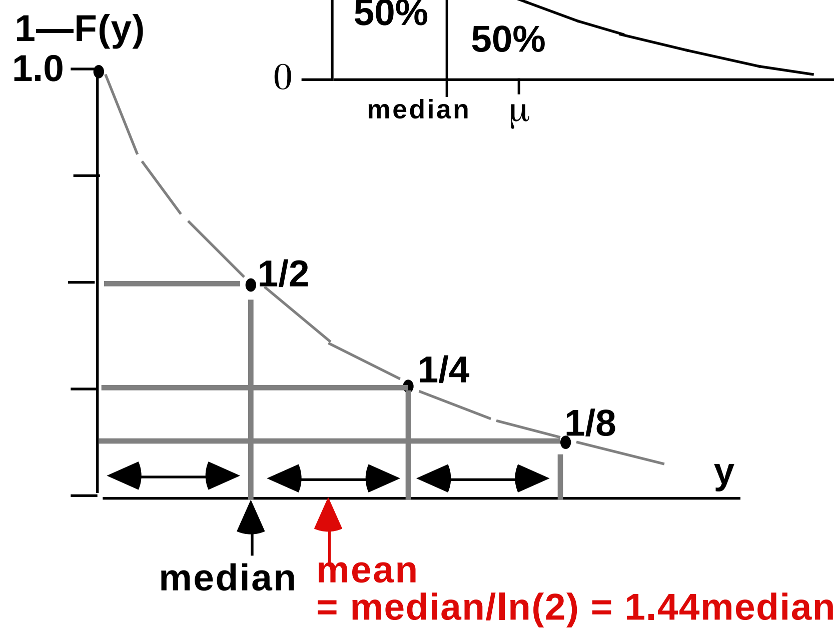
for the exponential distribution,
with mean E(Y) = μ = β ,

Var(Y) = the square of the mean,
SD(Y) = the mean of Y

Visually, the complement of the cdf = 1—F(y) = exp[—(1/μ) y]
and the pdf f(y) = (1/μ) exp[—(1/μ) y]

<< 50% (<u>??</u>%) of Y values are longer than the average
>> 50% (<u>??</u>%) of Y values are shorter than the average





The median is like the "halflife" in the exponential decay curve

Useful exercises...

- Understand the similarity of the shapes of the 1—F(y) and f(y) functions. the former is (1/μ) times higher than the latter. If μ is short (μ < 1), so that (1/μ) is > 1, the pdf is very much 'banked up' against the left (0) end of the (0,infinity) range of Y, and consequently less of the probability mass is to the right of any y value. Conversely, if μ > 1, ...

- Compare with cdf and pdf for 1st-ace and 1st-duplicate birthday probelms .. for these, the range of Y is finite, and the conditional probabilities are just that -- conditional.But the "decreasing pdf" is similar.

- Prove that [for <u>any</u> positive RV) the area under the 1—F(y) curve equals the mean μ.

- Practice going between the mean and the median (and the quartiles y_{25} and y_{75} etc.) of the exponential distribution

- Understand the link between the (Poisson) probability of observing a <u>number</u> y=0 of events for a fixed <u>time</u> <u>period</u> (0,T] and the distribution of the duration of inter-event intervals. Having no events in T units is the same as having the interval to the next event be greater than T.

  An inter-event interval of μ[interval] time units implies an "event rate" of 1/ μ[interval] events per time unit.

from text

- 4.55 (times between accidents)

- 4.54?? (I am a bit sceptical. What would it take for this distribution to fit? It would say that even if one has spent 10 hours on the job, and is not finished, one is 'no further ahead' than one who has spent 5 hours so far on another house. Maybe this "length of time" is <u>waiting</u> for the cement guy, or the plumber, who could show up at "any time", rather than doing any work oneself.

- 4.64, 4.66, 4.108, 4.109, 4.112

The Gamma distribution

A 2-parameter model: "scale" β >0; "shape" α > 0; Y≥0

WMS5 could have been more forthcoming about the "various reasons" (bottom of p 157) why the gamma distribution might be an appropraiete model for data, especially if they are of the "time until" variety.. And again, the parameters can be given more meaning.

Some insights can be gained from inspecting the expected value and variance of the gamma distribution, and how these relate to the corresponding ones for the exponential distribution.

	E(Y)	Var(Y)
Gamma	$\alpha\beta$	$\alpha\beta^2$
Exponential	α	β^2
Ratio	α	α

A first guess might be that the Gamma randomvariable Y_g is just a multiple α of the exponential random variable Y_e.
But this would not really fit, since if we simply multiply a single exponential random variable Y_e by a multiple α, the general shape of the distribution would not change, just its variance. And, moreover, the variance of ($\alpha \times Y_e$) would then be $\alpha^2 \times$ Var(Y_e), or $\alpha^2 \times \beta^2$, which is 'at variance' with the correct variance, namely $\alpha\beta^2$.

A second guess might be that the Gamma random variable Y_g is a sum of α DIFFERENT (and independent) exponential random variables (Y_e's) all with the same mean and varaince.

This fits better. We haven't dealt with variances of sums of independent RV's up to now, but it turns out that the variance of such a sum is the sum of their variances (as Mosteller tells his readers when motivating the concept of variance, and as will be introduced by WMS5 in section 8 of Chapter 5).

Var ($Y_{e1} + Y_{e2} + \ldots + Y_{e\alpha}$)

= Var (Y_{e1}) + Var(Y_{e2}) + ... + Var($Y_{e\alpha}$) = $\alpha \times$ Var(Y_e) = $\alpha \times \beta^2$,

as in Theorem 4.8. Thus if a process involves α independent Y_e's <u>in series</u>, then ΣY_e's ~ Gamma(α,β).

**Some notes on the gamma distribution,...
and some reflections on what might be giving rise to the distribution of the intervals between mammography and surgery for breast cancer surgery...**

(By the way, the gamma distribution is behind some inportant sampling distributios in Ch 7.2; definition 4.9 anticipates this)

If it takes α (α a positive integer) independent steps to gets to where one wishes, and each step involves the same exponential distribution of waiting time until someone "answers" or "does something", or the like, [average wait per step = β], then the sum of the α subintervals should have a gamma distribution with parameters α and β.

If $\alpha > 1$, the sum has a distribution with probability mass increasing until the mode (i.e., where the pdf is highest) at y=(α-1)b [CHECK!], and then falling away for y values past the mode. (see Fig 4.15 for examples with α=2 and α=6).

By the way, since β is just the mean for each component in the sum , one can w.l.o.g., take β=1 to help understand things.

Can we find values for α and β that would match the observed mean (approx 50) and SD (approx 42, so variance=1800) of the intervals in the 1998 breast cancer data...?

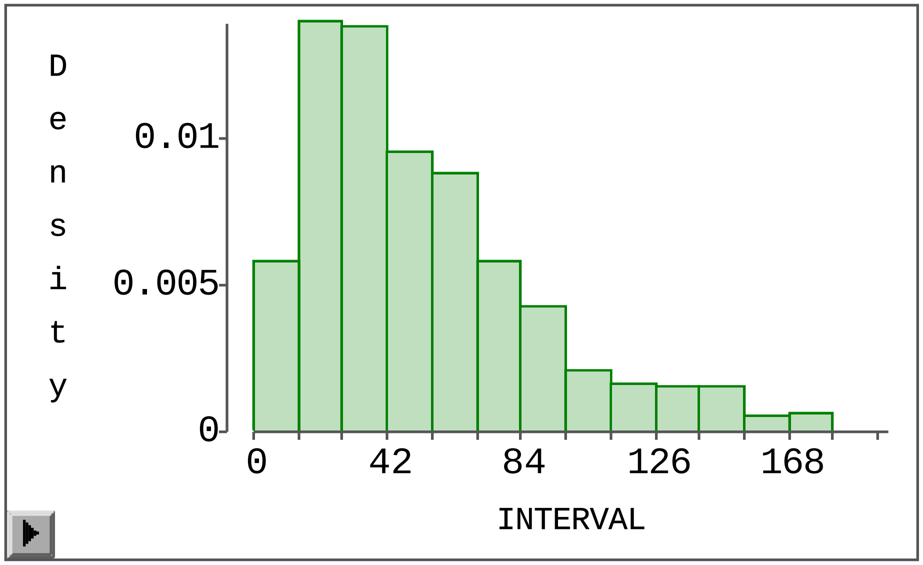
Distribution	(1) E(Y)	(2) SD(Y)	(3) Var(Y)	(4) (3)/(1)	(5) (1)/(4)
Gamma	$\alpha\beta$		$\alpha\beta^2$	β	α
1998 Data	55	42	1800	32.7	1.7

I haven't had time to superimpose the two curves!

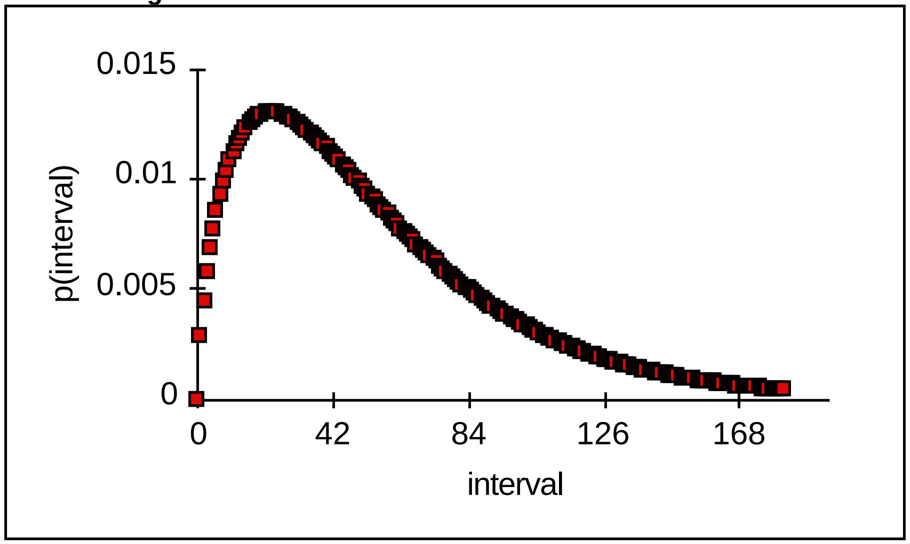
Of course, even if the theoretical model did fit well, that doesn't prove that this is how the intervals arose in real life.. (subintervals may not (i) have same distribution, (ii) be independent (iii) have exponential distributions, etc..)

Another way to think of the distribution: as a mixture of several non-exponential distributions.. some womem are in rural areas where the timimg is different, some are women who had larger tumors that may require fewer imaging tests, some patients' physicians get get OR time more easily, ...

Distribution of raw data from 1998 (n=500, from SAS INSIGHT)



Gamma distribution (pdf) with α=1.7, β = 32.7 computed using Excel GAMMADIST function



PS: Matching the mean & variance of the data & of the theoretical model the is an example of estimating parameters by the "Method of Moments" ($9.6)