1[1] Let Y be a random variable whose values are all between 0 and 10, and let F be its cumulative distribution function. What is the numerical value of F(11) — F(–1) ? Why?

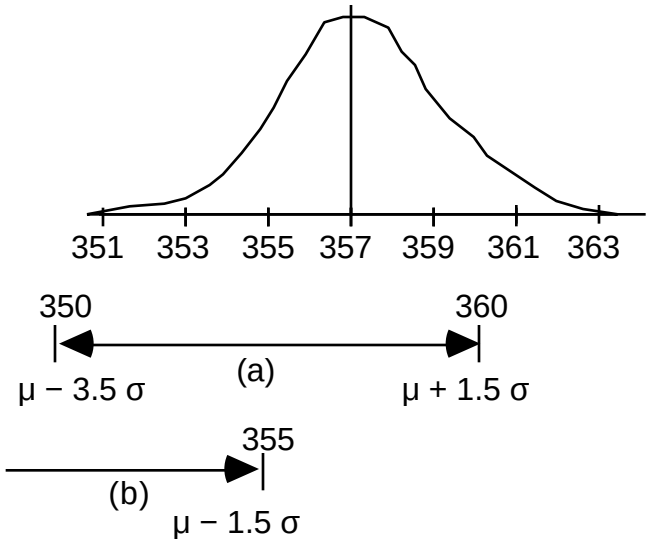
F(y) = Prob(Y ≤ y) , so F(-1) = 0 [no values below 0] and F(11) = 1 [none > 10]
So, F(11) - F(-1) = 1 - 0 = 1.

Somebody wrote " F(Y ≥ 10) = 1." What is wrong with this?

Someone else thought Y had to be U(0,10). But in fact F(-1) = 0 and F(11) = 1 no matter how the probability is cumulated over (0,10).

2[3] Soft drink cans are labelled as containing 355 ml of product. In fact, the automated canning process fills cans with varying amounts of liquid. Suppose that the actual fills of these soft drink cans are normally distributed with a mean of 357 ml and a standard deviation of 2 ml.

a What is the probability that a randomly selected can will contain (See Table on last page)
(i) between 350 ml and 360 ml (ii) less than the promised 355 ml , of soft drink?



(a) The percentage between 350 and 360 is the same as the percentage between Z = -3.5 and Z = +1.5, since 350 is (350-357)0/2 = 3.5, and (360-357)/2 = 1.5. Virtually 0% of probability is below μ-3.5σ, and some 6.7% is above μ+1.5σ. Therefore approx. 93.3% is between.

(b) 355 is 1SD below the mean; Because table reads the other way, we have to use the symmetry and fact that approx. 16% is above μ+1.5σ, so approx. 16% is below μ-1.5σ (or 355).

b What is the relation between the table and the distribution function F(z)?

The table gives Prob(Z > z), which is 1 - F(z)

3[3] The lifetime of a certain crucial component in a space shuttle follows an exponential distribution with a mean of 96 hours.

a If this space shuttle stays in space for three full days (i.e. 72 hours), what is the probability that this component will not have failed during this time? *Do not do the calculation: instead, show how you or your research assistant would do it, using first principles, or an Excel function, etc. It would help your assistant if you sketched a rough diagram!*

Y = Lifetime = How long it lasts = time until component fails or stops working.

Y ~ exponential[parameter β = 96 hours]; f(y) = $\frac{1}{\beta}$ exp[– $\frac{1}{\beta}$ y];

Q is about the prob. that component does not fail during the first 72 hours, i.e. that it does fail sometime after 72. This is the complement of F(72), i.e. 1 – F(72), or the area under the pdf(y) to the right of Y=72.
A number of you took the words "will not have failed during this time" to mean "within the first 72 hours". Perhaps there were too many negatives ("not", "fail" ..). It might have been better to ask about the probability that it does last longer than 72 hours.

It turns out that for the exponential distribution, both the pdf(y) and the cdf have closed form expressions. As given above,

pdf(y) = f(y) = $\frac{1}{\beta}$ exp[– $\frac{1}{\beta}$ y];
So cdf(y) = Prob(Y ≤ y) = Integral of f(t) from t= –Infinity (or zero) to t= y.

= 1 – exp[– $\frac{1}{\beta}$ y].
So, in our example with y=72 and b=96, cdf(72) = 1 - exp(– $\frac{72}{96}$) = 0.53,
or
Prob (Fail after 72) = 1 - cdf(72) = exp[– $\frac{1}{96}$ 72] = exp[-0.75] = 0.47

F(y) , or F(72) in this example, is also obtainable in Excel as

```
EXPONDIST(72,1/96,TRUE)
```

where 72 is the "y" value, 1/96 is what Microsoft call "lambda", and TRUE asks for the

cdf rather than the pdf.

All the textbooks I have been able to check say that the parameter (mu or theta or beta) of the exponential distribution is the mean lifetime or the mean time between failures if one that fails is immediately replaced. so we have the term $\frac{1}{\beta}$ in two places in the pdf. But what is the interpretation of $\frac{1}{\beta}$? If the average time to (or average time between) failures is 92 hours, that means that average number of failures in 96 hours is 1! [in order to count failures, we need to replace the component as soon as it fails, so that there is always one component in operation and thus 'at risk' of failing. This is why Microsoft (and I think I remember, some statistics books) reparametrize the exponentialpdf as

$\lambda \exp[-\lambda y]$

where now λ has the interpretation of "average number of events in a time unit, here 1 minute". With an average of β =96 hours between failures, the average number of failures per hour is indeed the small number 1/96.

Over 72 hours, you expect on average 72(1/96) = 3/4 = 0.75 failures. So we can work out the Poisson Probability of 0 (or 1 or 2 or whatever..) failures in 72 hours, as

$$\exp[-0.75]\frac{0.75^0}{0!} = \exp[-0.75] = 0.47$$

Note the two ways to get there

- via the continuous variable TIme to Failure and the chance that it will exceed the target time "72 or more" or

- via the discrete variable Number of Failures in the "0 to 72 hour window".

b For added security, it was decided to include a second of these components which would automatically begin operating at the instant the first component failed (if it did). That is, the first component would continue to operate until failure, at which time the second component would begin to operate. Suppose this second component also has an exponential lifetime with mean of 96 hours. On an extended trip lasting a full week (i.e. 168 hours), what is the probability that at least one of the components is still working at the end of the flight? (That is, that the total system of two components has not completely failed.) *Again, do not do the calculation, just show how you would do it .*

Now the target is "up for 192 hours of more" or "At most one Failure in the time window 0 to 192 hours. So, again, we have two ways to get there.. via a continuous random variable or via the Discrete variable Number of Failures

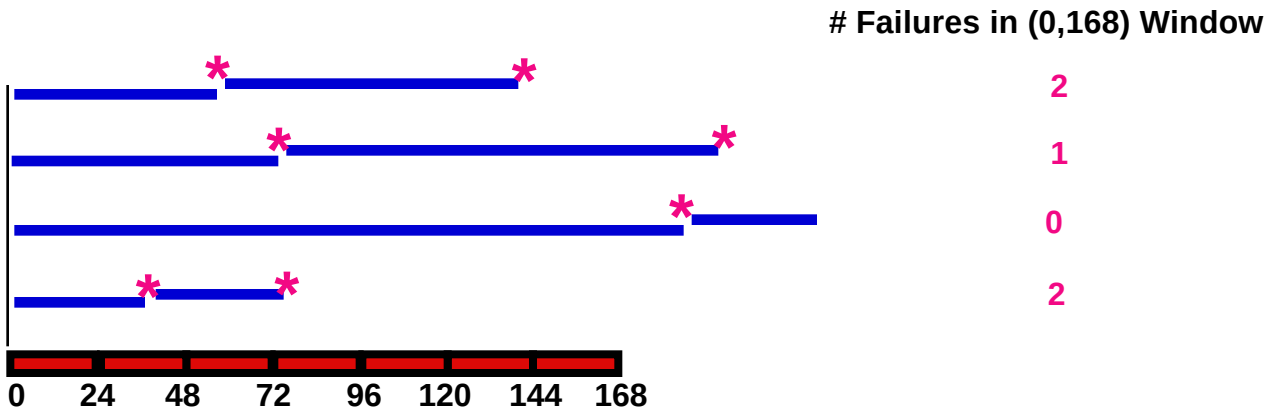
- via continuous variable route..

Think of the "total lifetime of the 2 components" we get by "stringing together" the 2 components.. as shown in blue below, The system succeeds if the total (potential or actual) up time is more than 168. The sum of the lengths of 2 (blue) lifetimes is Gamma with α=2 (because we added 2 independent exponentially distributed rv's) and β=96.

We would not have access to Excel in an exam, but

GAMMADIST(y=**168**,a=**2**,b=**96**,Cumulative=**TRUE**)

gives P(total ≤ 168) = 0.52, so that the probability of exceeding 168 is 48%.



- via Y = Number of Failures route..

Think of the number of failures we would have in the 168 hour window. The system succeeds if the total failures is 0 or 1. The average number of failures is 1 per 96 hours, or 168(1/96) = 1.75 per 168 hour window. The Poisson probability of y (=0,1,2, ..) failures in the window is given by

$$\exp[-1.75]\,\frac{1.75^{y}}{y!}$$

POISSON(y=**0**,mean=**1.75**,Cumulative=**FALSE**)= 0.173

POISSON(y=**1**,mean=**1.75**,Cumulative=**FALSE**)= 0.304

so prob(0 or 1 failures) = 0.477 = (to 2 decimal places) 0.48

Double check: POISSON(y=**1**,mean=**1.75**,Cumulative=**TRUE**)= **0.48.**

Key to getting this right is recognizing that there is only 1 component at risk at any one time, so the probability that this component fails in the next minute, given that it has been running for t minutes already, can be approximated by 1/96

(it is actually 1 - exp[-1/96], which is very close to 1/96, since exp[- small positive #] is approx. 1 - small positive #)

4[4] Suppose that response times for a computer to connect to the World Wide Web follow a gamma distribution with a mean of 6 seconds and standard deviation of $2\sqrt{3}$ seconds.

a Determine the parameters of the distribution of these response times, and sketch what the distribution looks like.

As a famous baseball manager Yogi Berra used to say[1], "this is déjà vu all over again". The first two moments (mean and variance) of the gamma distribution with parameters α and β are mean = $\alpha\beta$ = 6, variance = $\alpha\beta^2$. We are given the SD=$2\sqrt{3}$, so the variance is $\alpha\beta^2$ = 4 x 3 = 12. From these, we can back calculate that β = 2 and α = 3.

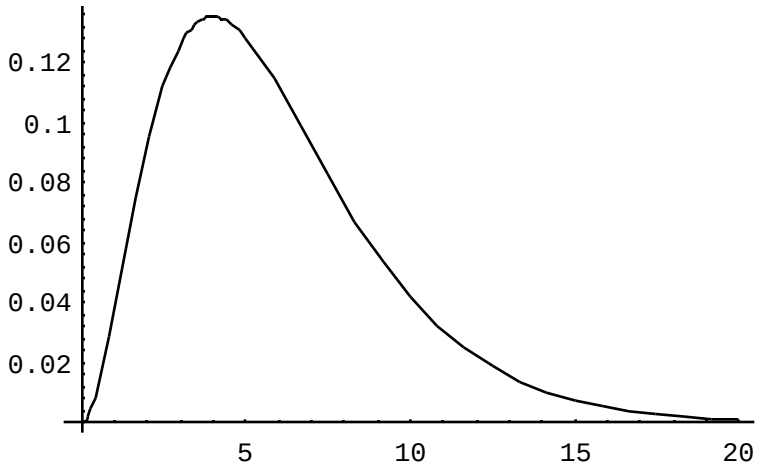
b Suppose that it has already been 5 seconds since you tried to connect, and that the computer has not yet connected. Show how to calculate the probability that the computer will connect in the next 5 seconds. *Be generic, i.e. do not make your method specific to the gamma distribution!*

WATCH THE WORDS HERE.. You are GIVEN that Y > 5.
If you ignore this, and just calculate the integral of the pdf [i.e., f(y)] function from y=5 to y=10, then you will get the <u>unconditional</u> probability that 5 ≤ Y ≤ 10.

If, instead, you <u>use the information</u> that Y≥5, then we are now talking about the conditional distribution of Y | Y ≥ 5. This is quite different from the unconditional one and the probability of interest is P(Y ≥ 10 | Y ≥ 5) = P(Y ≥ 10) / P(Y ≥ 5).

Pictorially...

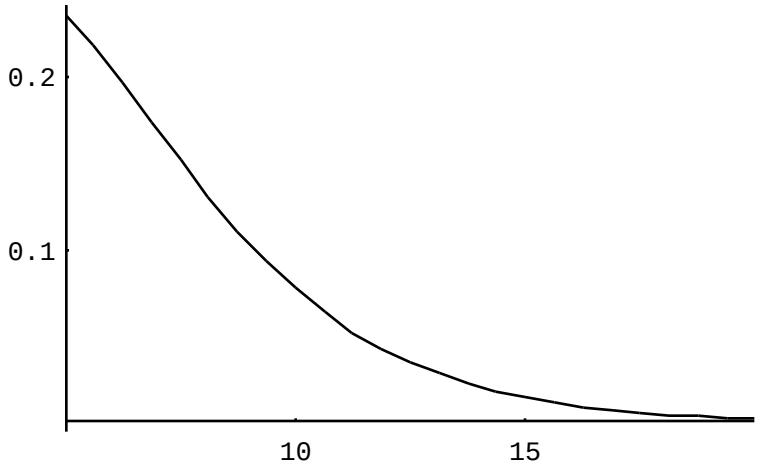
```
Plot[ PDF[GammaDistribution[3, 2], x],{x,0,20}]
```




```
CDF[GammaDistribution[3, 2], 5]

0.456187

CDF[GammaDistribution[3, 2], 10]

0.875348
```

[1] http://www.yogi-berra.com/yogiisms.html or http://www.yogiberraclassic.org/quotes.htm

```
CDF[GammaDistribution[3, 2], 10] - CDF[GammaDistribution[3, 2], 5]

0.419161

(CDF[GammaDistribution[3, 2], 10] - CDF[GammaDistribution[3, 2], 5]) /
(1 - CDF[GammaDistribution[3, 2], 5] )

0.770782

newpdf[y_]:= PDF[GammaDistribution[3, 2], y] /
             (1 - CDF[GammaDistribution[3, 2], 5] );

Plot[newpdf[y],{y,5,20}, PlotRange->{{5,20},Automatic},
Ticks->{{5,10,15},{.1,.2}}]
```




```
Integrate[newpdf[y],{y,5.,10.}]

0.770782
```

Some of you thought this was a case of the memory-less property. BUT this property is a property of the exponential only.. this is gamma with a=3 .. a bit more like Gaussian .. and we know that with Gaussian the conditional prob that Y is between "Yalready" and "Yalready + deltay" gets bigger as "Yalready" increases .. we discussed this when talking about terminating an interview (see my answers to exercise 4.112 from book (under solutions to selected problems from ch 4).

Only about 20% of you recognized the Conditional nature of the problem -- i.e. took notice of the fact that the chance of failure in the time 0-5 is now zero.. so if calculate the probabilities for all the intervals after 5, will come up short with the area (it will only add to P(Y>5). which is < 1. This is the same issue as calculating life expectancy from birth for persons whom we already know have survived 82 years!

Other issues: please be clear on labels for axes, and don't try to integrate the F(y) function.. it is already an integral.. you can integrate f to get F(y). and don't write about the prob(Y=y) when Y is continuous!

5[14] The following are data collected by Statistics Canada at the 1996 census:

Montreal Metropolitan Population by knowledge of official language (percentages are rounded, for sake of exercise)

Total	English only	French only	Both English and French	Neither English nor French
3,287,645	280,205	1,309,150	1,634,785	63,500
100%	8%	40%	50%	2%

a Consider a randomly chosen person. Let E be the variable denoting knowledge(Yes/No) of English, and F knowledge (Yes/No) of French. Use these data to display the joint distribution of E and F.

		Know French?		
		Yes	**No**	**Total**
Know English?	**Yes**	**0.50**	**0.08**	**0.58**
	No	**0.40**	**0.02**	**0.42**
	Total	**0.90**	**0.10**	**1.00**

b Use the data to create (i) the marginal distribution of E (ii) the marginal distribution of F.

Know English?	**Yes**	**0.58**
	No	**0.42**
	Total	**1.00**

Know French?		
Yes	**No**	**Total**
0.90	**0.10**	**1.00**

c Are E and F independent random variables?

NO. Various ways to test..

1. P(Know E and Know F) = 0.50 ≠ P(Know E) × P(Know F) = 0.58 × 0.90 = 0.522

2. If Independent, the marginal distribution for E, the conditional distribution of E given Know F, and the conditional distribution of E, given Do Not Know F, should all be the same. (see conditional ones in d below)

And likewise for the distributions of F, marginally and conditional on knowledge and non-knowledge, of E should be same as 2 conditional distributions

d Compare the probabilities that

(i) a person who knows French also knows English

		Know French?
		Yes
Know English?	Yes	0.50
	No	0.40
	Total	0.90

To make these into probabilities, we need to scale them so they add to 1.00 i.e., multiply both of them by 1.00/0.90), or (equivalently) divide by 0.90. i.e., 0.50/0.90 and 0.40/0.90, so required answer is 5/9 ths = 55%

Those of you who prefer formulae over visuals will see this as

$$\text{Prob(E=yes | F = yes)} = \frac{\text{P(E = yes and F=yes)}}{\text{Prob(F = yes)}} = \frac{0.50}{0.90}$$

(ii) a person who knows English also knows French

		Know French?		
		Yes	No	Total
Know English?	Yes	0.50	0.08	0.58

2 probabilities are 0.50/0.58 and 0.08/0.58, so required ans. is 50/58 = 86%

e Give **E** the numerical value 1 if the person knows English, and 0 if not. Likewise for **F**.

(i) What is the numerical value, and the meaning, of

E(**E**)? E(**F**)? E(**E**+**F**) *(also give a non-technical meaning)*

*(**bold** face for numerical r.v., regular face for expectation)*

V(**E**)? V(**F**)? V(**E**+**F**)

Expectation of E is 0 x Prob(E=0) + 1 x Prob(E=1) = 0 x 0.42 + 1 x 0.58 = 0.58

Likewise Expectation of F is 0.90.

Expectation of sum is Sum of Expectations = 0.58 + 0.90 = 1.48

Can also arrive at it by working out the values of the new rv E + F

Possible Values of **E** + **F**	Probabilities of these values	Value x Probability	Value^2 x Probability
0	0.02	0.00	0.00
1	0.48	0.48	0.48
2	0.50	1.00	2.00
Sum	1.00	1.48	2.48

prob(sum=1) = 0.40 + 0.08 = 0.48.

Meaning of expectations:

E(**E**): average value of the **E** variable in the population = proportion for whom **E**=1 = proportion who know English. (more accurate to say proportion for whom it is reported that they know English). Likewise the expectation E(**F**) = 0.90 is the proportion who know French.

E + **F** is the random variable obtained by summing .. in effect, it counts how many of the 2 languages a person knows. So E(**E**+**F**) = 1.48 is the average number of (these 2) languages people know (see table, where just less than 1/2 know 1 and 1/2 know 2 and a tiny fraction knows 0, so "average" of 1.48 makes sense as an average. Of course, just like saying that the average person has 1 testicle and 1 ovary, "average" is a numerical idea, and doesn't necessarily refer to any one "central" person. In contrast, the median value in a population does refer to a particular person.

Variances

var() = expected value of **E**^{2} + (expected value of **E**)^2 .

values of **E**^{2} are 0^2 and 1^2, with probabilities 0.42 and 0.58, so Expectation of **E**^{2} is 0 x 0.42 + 1 x 0.58 = 0.58. So var(**E**) = 0.58 - 0.58^2 = 0.58(1 – 0.58) = 0.2436

Likewise var(F) = 0.90(1 – 0.90) = 0.09. So a lot less variation w.r.t French, since almost all know it. Much lower certainty as to whether a person knows English.

(ii) Compute Cov(**E**,**F**) *[Use the Cov(**E**,**F**) = E(**EF**) — E(**E**)E(**F**) version]*

E(**EF**) = 0x0 x 0.02 + 0x1 x 0.40 + 1x0 x 0.08 + 1x1 x 0.50 = 0.50

covariance(**E**,**F**)

= average product minus product of averages

= 0.50 minus 0.58 x 0.90

= 0.50 minus 0.522

= -0.022

(iii) What is the relation ("< " or "=" or ">") between V(**E**+**F**) and V(**E**) + V(**F**) ?

Var(sum) = Var(**E**) + Var(**F**) + 2 x Covariance(**E**,**F**)

Since Covariance is negative, the Variance of the sum will be less than the sum of the variances.

Var(sum) = average (squared sum) - square of average sum

= $2.48 - 1.48^2 = 0.2896$

Check:

Var(sum) = Var(**E**) + Var(**F**) + 2 x Covariance(**E**,**F**)

= 0.2436 + 0.0900 + 2(-0.0220)

= 0.2896

f If you randomly chose 2 persons in Montreal, what is the probability that they could understand each other in at least one of the two official languages (sign language and Franglais not allowed!)?

Hint: Label them Person1 and Person2, and use the data at the top of page.

MAKE A 2-way table... and determine (with an X) which sample points fir the definition of the "event", and the joint probabilities of these (obtained by simple multiplication of marginal probabilities, since person 1 selected independently of person 2.

		Person 2			
		E and F (0.50)	E only (0.08)	F only (0.40)	neither (0.02)
Person 1	E and F (0.5)	X 0.2500	X 0.0400	X 0.2000	-
	E only (0.08)	X 0.0400	X 0.0064	-	-
	F only (0.40)	X 0.2000	-	X 0.1600	-
	neither (0.02)	-	-	-	-

The probabilities can also be obtained from the expansion of $(0.50 + 0.08 + 0.42 + 0.02)^2$.

The probability of interest is the sum of the probabilities associated with the 7 sample points that meet the criterion, namely 0.8964, or close to 90%.

Probability Distributions

Distribution	Probability Function f(y)	Mean	Variance
Uniform	$\frac{1}{\text{upper} - \text{lower}}$	$\frac{\text{upper} + \text{lower}}{2}$	$\frac{[\text{upper} - \text{lower}]^2}{12}$
Exponential	$\frac{1}{\beta} \exp[-\frac{1}{\beta} y]$	β	β^2
Gamma	$\frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} \exp[-\frac{1}{\beta} y]$	$\alpha\beta$	$\alpha\beta^2$
Gaussian ("Normal")	$\frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{[y-\mu]^2}{2\sigma^2}]$	μ	σ^2

***Excerpts from Table on inside front cover of WMS5:-**

Normal curve areas: standard normal probability (prob) in **right-hand tail**:

z:	-1.0	-0.5	0	0.5	1.0	1.5	2.0	2.5	3.0
prob.:	0.841	0.691	0.500	0.309	0.159	0.067	0.023	0.006	0.001