WHY IS THIS CHAPTER IMPORTANT?

Virtually all statistical inference is based on statistics calculated using data from samples. The most common statistics are means, proportions, correlations, regression coefficients etc..

Take one of the simplest cases: say we are interested in estimating the "population" mean μ from sample observations $y_1, y_2, \ldots y_n$.

Here the relevant statistic is $\bar{y} = \frac{1}{n}y_1 + \frac{1}{n}y_2, \ldots + \frac{1}{n}y_n$,

a linear combination of the n random variables Y_1 to Y_n.

But to estimate its "goodness", one must also use the variability of the n values, measured by a statistic that involves computing

$$s^2 = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \ldots + (y_n - \bar{y})^2}{n-1}$$

which, apart from the divisor, is a sum of the squares of functions of the n random variables Y_1 to Y_n.

In fact, a key item in estimating (or testing) the mean μ is the statistic

$\bar{y} \pm$ some multiple of s,

with the "multiple" chosen so that the interval has a certain "coverage".

UP TO NOW, WE HAVE ONLY LEARNED HOW TO GET THE EXPECTED VALUE (MEAN) OF A FUNCTION OF ONE OR MORE RANDOM VARIABLES. IF WE ARE TO ASSESS VARIATIONS FULLY, WE NEED TO BE ABLE TO CHARACTERIZE THE FULL DISTRIBUTION OF WHATEVER FUNCTION OF THE n RANDOM VARIABLES IS OF INTEREST, NOT JUST THE MEAN (or the mean AND the variance) OF THE DISTRIBUTION!

Thus, we need to be able to be able to..

(1) given a full characterization of a random variable Y in terms of either f(y) or F(y), fully characterize the distribution of a function U of Y.

(2) given a full characterization of the [in a simple random sample] independent and identically distributed random variables RV_1 to RV_n, fully characterize the distribution of their sum, Sum = $RV_1 + \ldots + RV_n$.

Of course, if we can get the distribution of the sum, then the distribution of the mean is just a scaled version of the distribution of the sum.

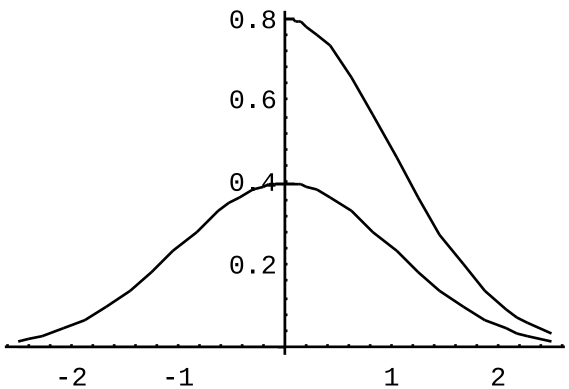
Chapter 7 gives the distributions for a number of the common statistics -- functions of y_1 to y_n.

But, how far could we get on our own...? What if we were curious enough to ask...

1a	If Y had a <u>Gaussian</u> distribution (with, say, $\mu = 0, \sigma = 1$), what would be the distribution of .. (i) $\lvert Y \rvert$? (ii) Y^2 ? ...

For 1a (i), we can imagine "folding"' the distribution over onto the positive axes.. so $pdf_{|Y|}(\bullet) = 2\ pdf_Y(\bullet)$ for $\bullet \geq 0$

```
f[y_]:=(1/Sqrt[2Pi]) Exp[-0.5 y^2];

Plot[{f[y],If[y<0,0,f[y]+f[y]]},
  {y,-2.5,2.5},
  AxesLabel->{"y or abs[y]","pdf(Y) and pdf[abs(Y]"}]
```



For (1b) we can imagine "mapping" the Y values into $X = Y^2$ values... and keeping track of the probabilities in each little sub-region.. e.g.,

Since..		of the Y's are between	This same percentage of the $X = Y^2$ values will be between...
	50%	-0.67 and +0.67,	0 and 0.67^2 = 0.45
	68%	-1 and +1,	0 and 1
	80%	-1.28 and +1.28,	0 and 1.28^2 = 1.64
	90%	-1.645 and 1.645	0 and 1.645^2 = 2.71
	95%	-1.96 and 1.96	0 and 1.96^2 = 3.84
	etc		

But it is labour intensive and doesn't produce an analytic expression for the pdf or cdf of the RV Y^2. [Of course, we are not using the analytic expression for the cdf or pdf of Y.]

Another example..

1b	If Y had a <u>Uniform</u> [say on the interval(0,1)], so $f_Y(y)=1$ on (0,1), what would be the distribution of .. (i) $X = Y^2$ (ii) $X = Y^{1/2}$? (iii) $X = \ln(Y)$? ...

Since..		of the Y's are between	this same percentage of the $X = Y^2$ values will be between...
	10%	0 and 0.1,	0 and 0.1^2 = 0.01
	10%	0.1 and 0.2,	0.01 and 0.04
	..	etc	etc
or, Since..		of the Y's are below	this same percentage of the $X = Y^2$ values will be below...
	10%	0.1	*0.01*
	20%	0.2	*0.04*
	etc		
Better still..†		of the Y's are below	this same percentage of the $X = Y^2$ values will be below...
	10.0%	$0.01^{1/2} = 0.100$	*0.01*
	14.1%	$0.02^{1/2} = 0.141..$	*0.02*
	17.3%	$0.03^{1/2} = 0.173..$	*0.03*
	20.0%	$0.04^{1/2} = 0.200$	*0.04*
	100% × $x^{1/2}$	$x^{1/2}$	any value, x
			etc

†start with rightmost column, and work left...

This last way allows us to get the cdf for $X = Y^2$ directly, in <u>linear</u> increments of Y^2. If we now want the pdf for the RV $X=Y^2$, then we can take the derivative (or use successive subtractions)

We reasoned that the proportion of the probability distribution of X that is to the left of X = x is given by:

$$\text{Prob}(X \le x) = x^{1/2}.$$

i.e

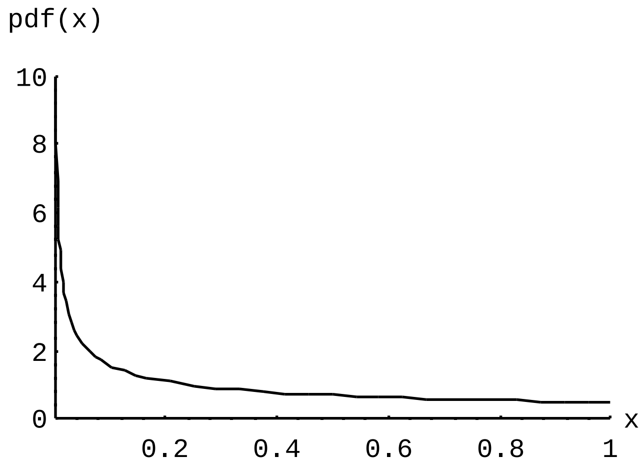
$$F_X(x) = x^{1/2}.$$

So

$$f_X(x) = \frac{dF_X(x)}{dx} = (1/2)\, x^{-1/2}.$$

Check: is this a pdf? does it integrate to 1?

```
g[x_]:= (1/2) x^(-0.5);
Plot[g[x],{x,0,1},
PlotRange->{{0,1},{0,10}},
AxesOrigin->{0,0},
AxesLabel->{"x","pdf(x)"}]

Integrate[g[x],{x,0,1}]

1.
```



(ii) Probability Distribution of the r.v. $X = Y^{1/2}$, [Y ~U(0,1)]

Prob $(X \le x)$ = Prob $(Y^{1/2} \le x)$ = Prob $(Y \le x^2) = x^2 = F_X(x)$

So $\quad f_X(x) = \dfrac{dF_X(x)}{dx} = 2x.$

This makes sense, since more of the X values will now be towards the X=1 end of the (0,1) scale, and the integral of $f_X(x) = 2x$ over (0,1) is 1.

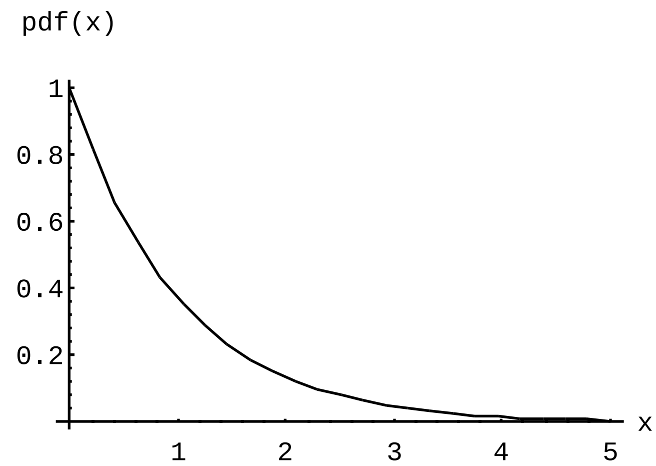
(iii) It is the same procedure if we created the rv X = –ln(Y).

$F_X(x)$ = Prob $(-\ln(Y) \le x)$ = Prob $(Y \ge e^{-x}) = 1 - e^{-x}$

So $\quad f_X(x) = \dfrac{dF_X(x)}{dx} = (-)e^{-x}(-1) = e^{-x}$

Again, this makes sense, since more of the X = -ln(Y) values will now be concentrated towards the X=0 end of the (0,Infinity) scale, and the integral of $f_X(x)$ over (0,Infinity) is 1.

```
g[x_]:= Exp[-x];
Plot[g[x],{x,0,5},
PlotRange->All, AxesOrigin->{0,0},
AxesLabel->{"x","pdf(x)"}]
```




```
Integrate[g[x],{x,0,Infinity}]
1
```

TAKE-HOME MESSAGE

If the cdf of the original (continuous) variable Y is tractable, then the above **Method of Distribution Functions** (cdf) route is a good way of arriving at the cdf (and thus pdf) of the new random variable. The book also suggests it also for a function of n random variables, but that may involve some serious n-dimensional integration!

Are there even more direct ways, especially if cdf is not tractable (e.g. Gaussian) but the pdf is?

Two examples

1c

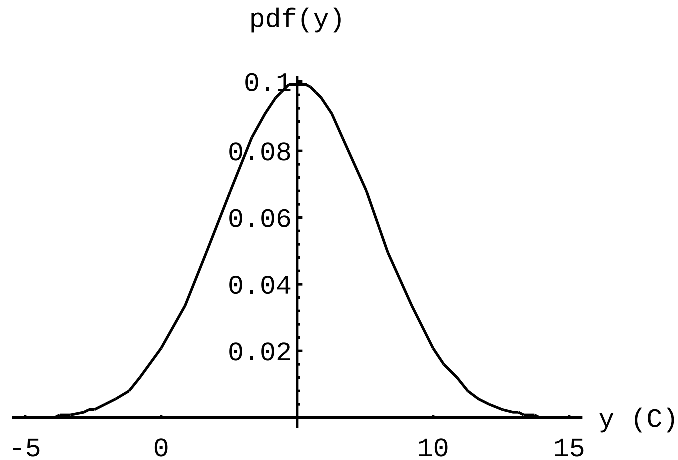
(i)	(déjà vu) If Y has a Uniform distribution on say (0,1), what is the distribution of the random variable $X = Y^2$?

[N(0,1) is shorthand for "Normal with mean 0, SD 1"]

(ii)	If Y = temperatures (C) in Montreal in a certain month have a Gaussian distribution (with, say, $\mu = 5, \sigma = 4$), what would be the distribution of X = the temperatures in F?

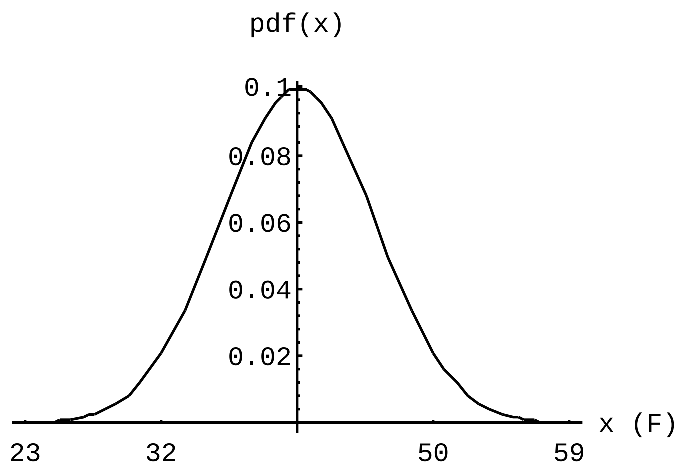
Let's start with easier one, (ii), where we know the answer.

```
meanC = 5; meanF = (9/5)meanC + 32;

sdC=4; sdF = (9/5) sdC;

pdfY[y_]:=(1/(sdC Sqrt[2Pi])) *
          Exp[-((y-meanC)/sdC)^2];

pdfX[x_]:=(1/(sdF Sqrt[2Pi])) *
       Exp[-((x-meanF)/sdF)^2];

Plot[pdfY[y],{y, meanC - 2.5sdC, meanC + 2.5sdC},
AxesOrigin->{meanC,0}]
```



To get the distribution in Farenheit, what would be wrong with taking the above graph and just changing the C values on the horizontal axis to their equivalents in Fahrenheit?

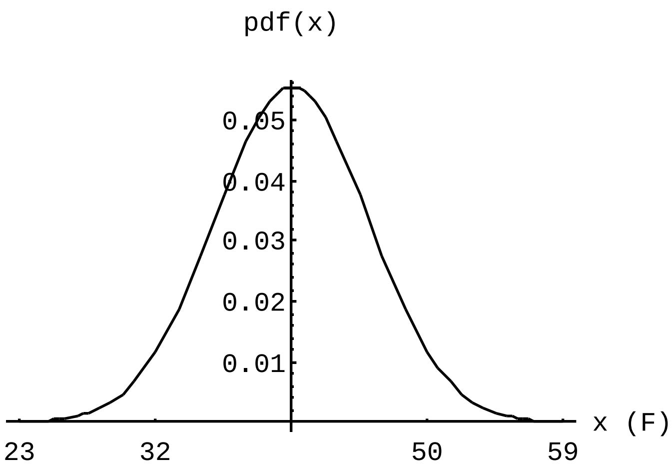
```
Plot[pdfY[y],{y, meanC - 2.5sdC, meanC + 2.5sdC},
AxesOrigin->{meanC,0}, AxesLabel->{"x (F)","pdf(x)"},
Ticks->{ {{-5,"23"}, { 0,"32"}, { 5,"41"},  {10,"50"},
         {15,"59"} },   Automatic} ]
```



Answer: The area under the curve is now >> 1 !

If we enlarge the horizontal scale by a factor of 9/5, we must shrink the height of the pdf so that the area is unity. i.e.

```
Plot[(5/9)pdfY[y],{y, meanC-2.5sdC, meanC + 2.5sdC},
AxesOrigin->{meanC,0}, AxesLabel->{"x (F)","pdf(x)"},
Ticks->{ {{-5,"23"}, { 0,"32"}, { 5,"41"}, {10,"50"},
          {15,"59"} },  Automatic}  ]
```



You can see by your eye that the area under this curve is more appropriate (it does in fact integrate to unity).

In this very simple (linear transformation) example, this translates to the following rule:

pdf for new random variable X (X = fn. of old rv Y)

pdfX(x) = pdfY("y-equivalent of x") × scale factor,

where scale factor is the ratio of the 2 scales..

ratio = $\dfrac{\textbf{old scale (C in e.g.)}}{\textbf{new scale (F in e.g.)}}$

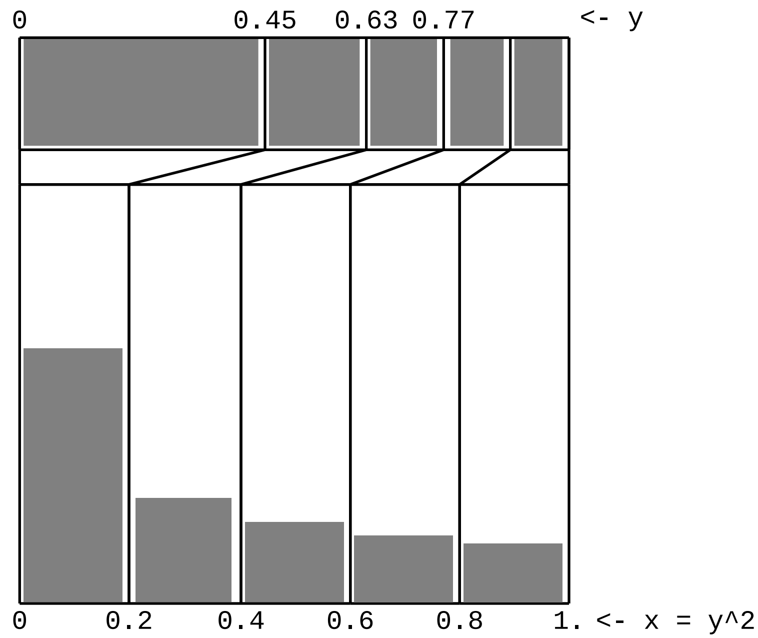
< 1 if new scale (X) is larger than old one (Y)

> 1 if new scale (X) is smaller than old one(Y).

It IS NOT SO MUCH THAT WE CREATED A NEW RANDOM VARIABLE X FROM THE OLD ONE Y AS IT IS THAT WE CHANGED THE SCALE ON WHICH THE VARIABLE WAS MEASURED. CONCEPTUALLY, THERE IS ONLY ONE RANDOM VARIABLE TEMPERATURE. HOW WE MEASURE IT CHANGES THE SCALE.

Now, let's do the harder one, (i), where the change of scale depends on where on the new scale one is...

(i)	(déjà vu) If Y has a Uniform distribution on say (0,1), what is the distribution of the random variable $X = Y^2$?



Think of "pouring" the probability mass from the original (old) containers (top) into the new ones at bottom, ALL THE WHILE CONSERVING THE 100% PROBABILITY MASS I.E. THE AR EA UNDER THE PDF FOR THE NEW R.V. MUST EQUAL 1, JUST AS THE AREA UNDER THE ORIGINAL PDF EQUALS 1.

NOTE THAT THE AREA OF EACH RECTANGLE SITTING ON THE NEW SCALE (BOTTOM) MUST EQUAL THE AREA OF THE CORRESPONDING RECTANGLE IN THE ORIGINAL DISTRIBUTION (TOP). SO IF WE BASES OF THESE RECTANGLES ARE OF UNEQUAL SIZES, THEIR HEIGHTS MUST BE ALTERED ACCORDINGLY SO AS TO KEEP THE TWO AREAS EQUAL.

If we make the containers (or "bins") narrower, the density curve (the heights of the containers) becomes smoother.

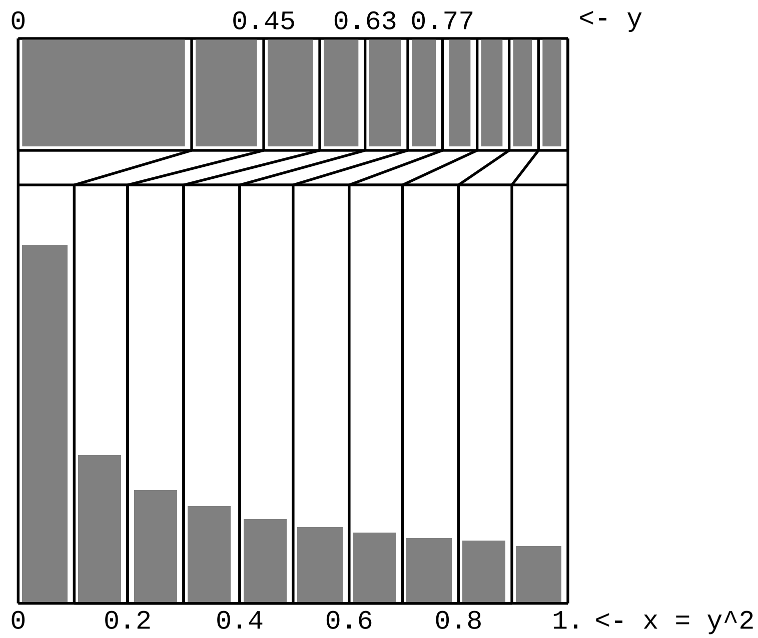
height of lower container

= height of upper container × scale factor

where scale factor is the LOCAL ratio of the 2 scales..

$$\text{ratio} = \frac{\text{old scale at point}}{\text{new scale at point}} = \frac{dY}{dX}$$

< 1 if new scale (X) is changing faster than old one (Y)
> 1 if new scale (X) is changing "slower" than old one(Y).



Cf. exact version: $f_X(x) = (1/2)\,x^{-1/2}$.

Note that there is nothing in the above that says that the new random variable must have the same range as the old one. For example it would work fine if say

Y ~ Uniform	**on (0.1)**
X = -ln(Y)	**on (0,Infinity)**

In this case, X is a decreasing function, so it is more difficult to draw my "pouring probability from containers" diagram.

pdf for X

$pdf_X(x) = pdf_Y(\text{"y-equivalent of x"}) \times$ **scale factor,**

where scale factor is the (always positive) local ratio of the 2 scales.. i.e.,

$$\left| \frac{dY}{dX} \right|$$

$X = -\ln(Y)$

==> $Y = e^{-X}$

==> $\frac{dY}{dX} = (-1)e^{-X}$

==> $\left| \frac{dY}{dX} \right| = e^{-X}$

==> $pdf_X(x) = pdf_Y(\text{"y-equivalent of x"}) \times e^{-x}$

$= pdf_Y(e^{-X}) \times e^{-x}$

But since Y ~ Uniform on (0,1), $pdf_Y(\bullet) = 1$ on (0,1), we get

$pdf_X(x) = 1 \times e^{-x} = e^{-x}$

Note that this Method of Transformations of finding the pdf of a function of a continuous random variable only works if the function X of Y is decreasing or increasing. For example, it will not work if say

$Y \sim N(0,1)$ and $X = Y^2$

On re-reading it, I find WMS5's explanation of the Method of Transformations quite short and clear -- provided you recognize the <u>Chain Rule</u> in the last step.. (They use U for the new variable; I tried to stay away from U because it also denotes Uniform").

For X and increasing function case..

$P(U \le u) = P(Y < y)$, where y -s the y-equivalent of u

i.e.

$F_U(u) = F_Y(y)$

Thus,

$$f_U(u) = \frac{dF_U(u)}{du} = \frac{dF_Y(y)}{du} = \frac{dF_Y(y)}{dy}\,\frac{dy}{du}$$

$$= pdf_{OLD}(\text{evaluated at y-equivalent of u}) \times \frac{d\text{ "old"}}{d\text{ "new"}}$$

For X a decreasing function case..

notice that now

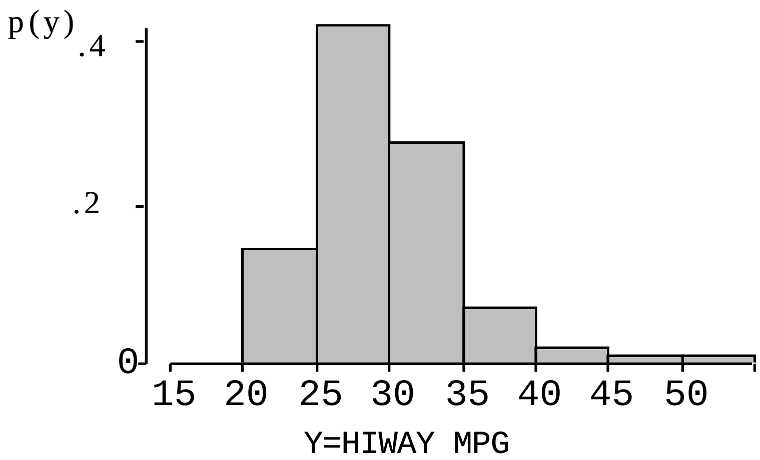
$P(U \le u) = 1 - P(Y < y)$,

See book p266 -267 for why no matter which, the scale factor is the absolute value of $\frac{d\text{ "old"}}{d\text{ "new"}}$.

TAKE-HOME MESSAGE

If the pdf of the original (continuous) variable Y is tractable, and the transformation is monotonic, then the above **Method of Transformations** (cdf) route is a good way of arriving at the pdf of the new random variable.

Example Y = fuel economy, measured in miles per gallon (mpg) on the highway, of 1993 model automobiles..



Q: What is the probability histogram for

X = the number of gallons of gas to go 300 miles?

i.e X = 350 / Y

AGAIN, THE KEY IS TO KEEP THE TOTAL AREA UNDER THE NEW HISTOGRAM AS UNITY, SO IF YOU USE A BASE FOR A RECTANGLE THAT IS SMALLER (LARGER) THAN THAT OF THE CORRESPONDING RECTANGLE IN THE ORIGINAL DISTRIBUTION, YOU NEED TO MAKE THE HEIGHT OF THE NEW RECTANGLE LARGER (SMALLER) SO AS TO MAINTAIN THE SAME RELATIVE FREQUENCY FOR THE NEW AND OLD RECTANGLES.

Example Given:

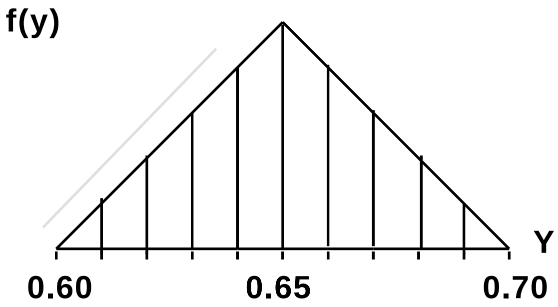
Y = What Canadian $ is worth, in $US over the last year ($pdf_Y(y)$ say)

Q:

X = what a US$ is worth, in $ Canadian, in same period?

X = 1/Y

Say distribution (pdf) of Y is given by..



Note that this distribution uses 10 "equal-width bins".

Q: Say we want to also use 10 equal width "bins" for the distribution of X = 1/Y

The range of X is 1/0.7 = 1.43 to 1.67, a total distance of 24 cents], so each bin will be 2.4 cents wide.

So the heights in the X bins will have to be a lot lower so as to make the area in the rectangle, and the overall area, for X come out to unity.

WHAT ARE THEY AND WHAT IS THEIR MAIN USE?

Remember we tried to "match up"

> **an empirical distribution (Breast Cancer waiting times) and a theoretical distribution (Gamma).**
>
> **in the quiz, another empirical distribution (waiting times to connect to the WWW) and a theoretical distribution (Gamma).**

In both instances, we did so by matching the means and the variances. But one could have two non-identical distributions that had the same mean and variance.. they could differ somewhere along their pdf's or cdf's.

What if we matched them not just on the average squared deviations from the mean, but also the cubed deviations, the deviations to the 4th power, etc..?

Ultimately, if the distributions matched on all these characteristics for as high a power as we cared to go, then we would say they were identical (a bit like with fingerprints) See the uniqueness Theorem at bottom of p 271. Thus, they help us to "recognize" certain distributions.

For example, it could take some work to prove -- from scratch -- that the sum of 2 or more independent Gaussian random variables is again Gaussian, but since we can work out the "signature" moment generating function for the Gaussian distribution, and since the moment generating function of a sum of i.i.d rv's is simply the product of their respective moment generating functions, we can quickly recognize that the sum must also have a Gaussian distribution.

What is fast way to get all these "moments" (k-th moment is average of k-th power of deviation from mean -- can also calculate the deviations from zero, but they are more awkward)?

The trick is to use the expected value of an infinite series in some argument "t" ..

i.e. if Y is a random variable, we define its moment generating function, $m_Y(t)$ as

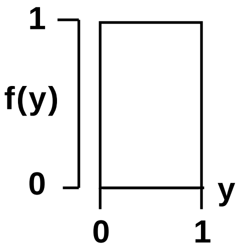
$$E(e^{Yt})$$

where E denotes the expected value, e is the base for the natural logs and t is the "argument" for the moment generating function. (see p 119)

But e to a power of Yt is an infinite series, involving powers of Y, and thus $m_Y(t)$ involves all of the moments of Y. One can extract these moments by repeated differentiation and subtraction.

Random Variable	**Sum of 2 i.i.d. such variables**	**Sum of 12 i.i.d. such variables**	**Sum of 30**

Uniform on (min,max)

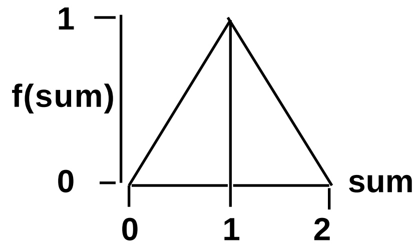


$f(y) = \frac{1}{\text{min— max}}$ eg min 0, max 1

$\text{mean} = \frac{\text{min+max}}{2}$ $\text{Var} = \frac{[\text{max-min}]^2}{12}$

$\frac{1}{2}$ $\frac{2}{12}$

e.g. min=0, max=1

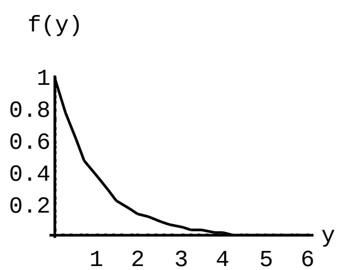


mean = 1 $\text{Var} = \frac{2}{12}$

Exponential on (0,Infinity)

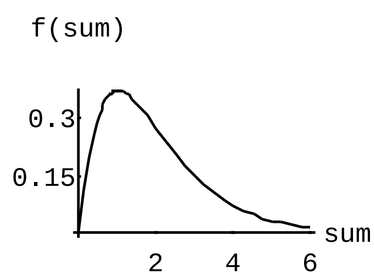
$f(y) = \frac{1}{\beta} \exp[— \frac{1}{\beta} y]$

mean = β (1 in e.g. below) ; var = β^2



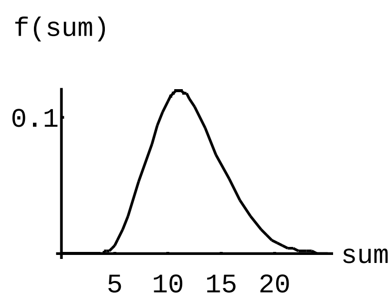
Gamma (α = 2, β[= 1 here] **)**

mean = 2β ; var = $2\beta^2$



Gamma (α = 12, β [= 1 here] **)**

mean = 12β ; var = $12\beta^2$



Gamma on (0,Infinity)

$$\frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{\alpha-1} \exp[-\frac{1}{\beta} y]$$

mean = αβ; var = $\alpha\beta^2$

Normal (μ, σ) on (-Inf,Inf)

f(y) =

$$\frac{1}{\sigma\sqrt{2\pi}} \exp[- \frac{[y-\mu]^2}{2\sigma^2}]$$

Gamma on (0,Infinity)

$$\frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{\alpha-1} \exp[-\frac{1}{\beta} y]$$

mean = 2αβ; var = $2\alpha\beta^2$

sum ~ Normal (2μ, $\sigma\sqrt{2}$)

$\bar{y}$ ~ **Normal** $(\mu, \frac{\sigma}{\sqrt{2}})$

$$\frac{\bar{y} - \mu}{\sigma / \sqrt{2}} \sim \textbf{Normal (0,1)}$$

$$\frac{\bar{y} - \mu}{s / \sqrt{2}} \sim$$ **Student's t (**1 df **)**

Gamma on (0,Infinity)

$$\frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{\alpha-1} \exp[-\frac{1}{\beta} y]$$

mean = 12αβ; var = $12\alpha\beta^2$

sum ~ Normal (12μ, $\sigma\sqrt{12}$)

$\bar{y}$ ~ **Normal** $(\mu, \frac{\sigma}{\sqrt{12}})$

$$\frac{\bar{y} - \mu}{\sigma / \sqrt{12}} \sim \textbf{Normal (0,1)}$$

$$\frac{\bar{y} - \mu}{s / \sqrt{12}} \sim$$ **Student's t (**11 df **)**

sum ~ Normal (30μ, $\sigma\sqrt{30}$)

$\bar{y}$ ~ **Normal** $(\mu, \frac{\sigma}{\sqrt{30}})$

$$\frac{\bar{y} - \mu}{\sigma / \sqrt{30}} \sim \textbf{Normal (0,1)}$$

$$\frac{\bar{y} - \mu}{s / \sqrt{30}} \sim \textbf{Student's } t_{29}$$