

1 Density functions, cumulative density functions, measures of central tendency, and measures of dispersion

densityfunctions-intro.tex October 1, 2009

Note that this section of notes is limited to the consideration of a single and continuous random variable.

Remember from earlier that the variable X is a one-dimensional, continuous random variable if there exists a function $f(x)$ such that $f(x) \geq 0 \forall x$ in the interval $-\infty \leq x \leq \infty$, and the probability that $(a \leq x \leq b)$ is

$$\Pr(a \leq x \leq b) = \int_a^b f(x)dx$$

The function $f(x)$ is called a *density function* (or a *probability density function*).

1.1 Examples of density functions

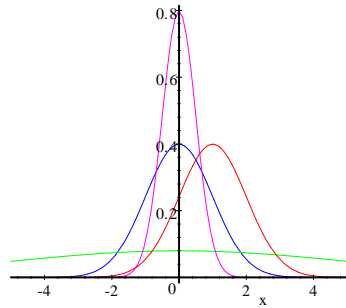
the normal:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-u)^2/2\sigma^2}$$

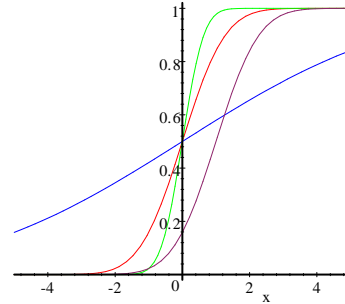
is a well known density function where σ and u are parameters in the density function.

Who first expressed this formula? Who first used it as a density function?

Graphs of the normal probability density functions are the familiar bell-shaped curves. The following plots show the probability density functions $\text{NormalDen}(x; \mu, \sigma)$, cumulative distribution functions $\text{NormalDist}(x; \mu, \sigma)$ for the parameters $(\mu, \sigma) = (0, 1), (0, 5), (0, 0.5), (1, 1)$.



Normal density functions



Normal distribution functions

But, be warned that $\int_{-\infty}^t \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-u)^2/2\sigma^2} dt$ does not have a closed-form solution.¹

Lets start more simply.

¹Note that $\sqrt{2\pi}\sigma = \sqrt{2\pi\sigma^2}$, ones sees the density function written both ways, and the distinction is easy to miss.

1.1.1 Consider the Following Four Density Functions

Example 1 Assume

$$g(x) = 0 \text{ if } x < 0 \text{ or } x > 3$$

and

$$g(x) = 3x, \quad 0 \leq x \leq 3$$
$$\int_0^3 3x dx = 1.5x^2 \Big|_0^3 = 1.5(9) = 13.5 \neq 1$$

So, $g(x)$ is not a density function, but

$$f(x) = \frac{3x}{13.5}$$

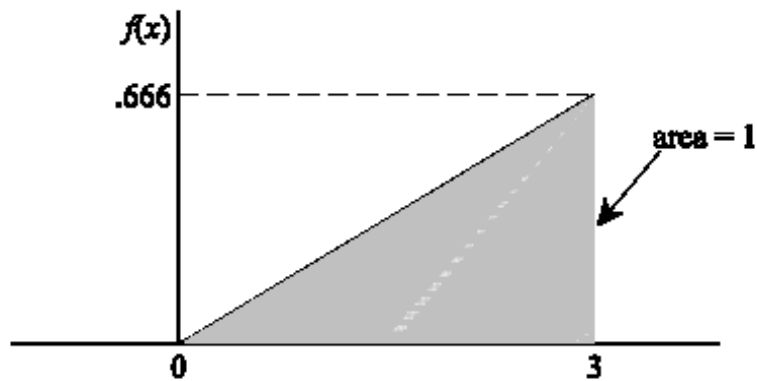
is.

That is

$$f(x) = 0 \text{ if } x < 0 \text{ or } x > 3$$

and

$$f(x) = .222x \text{ if } 0 \leq x \leq 3$$



Note the importance of being explicit about where $f(x) = 0$.

Example 2

$$f(x) = \left(\frac{1}{18}\right)(3 + 2x) \quad 2 \leq x \leq 4$$
$$f(x) = 0 \quad \text{otherwise}$$

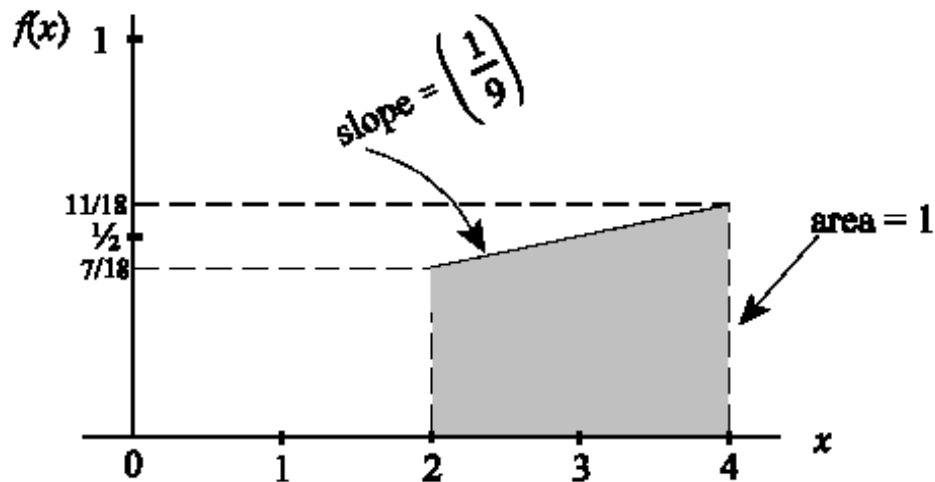
Is this a density function?

$$f(x) \geq 0 \quad \forall x$$

Integrate it to see if the area under it is one.

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-\infty}^2 0 dx + \int_2^4 \left(\frac{1}{18}\right)(3 + 2x) dx + \int_4^{\infty} 0 dx \\ &= \frac{1}{18} \int_2^4 (3 + 2x) dx = \frac{1}{18} (3x + x^2) \Big|_2^4 = \frac{1}{18} [(12 + 16) - (6 + 4)] \\ &= 1 \end{aligned}$$

So, yes.



Example 3

$$f(x) = se^{-s(x-n)} \exp[-e^{-s(x-n)}]$$
$$s > 0$$

The parameters are s and n .

Use *Mathematica* to graph this function and see how it changes as s and n change.

$$f(x) \geq 0 \quad \forall x$$
$$\int f(x)dx = \int se^{-s(x-n)} \exp[-e^{-s(x-n)}] dx$$
$$= \exp[-e^{-s(x-n)}]$$

Why? Recollect that

$$\int m'(x)e^{m(x)} dx = e^{m(x)}$$

So

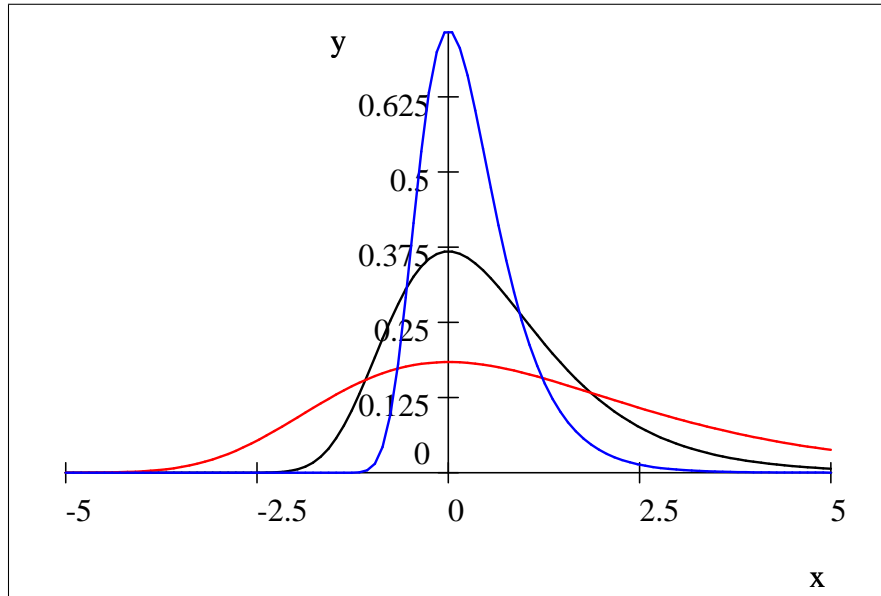
$$\int_{-\infty}^{\infty} f(x)dx = \lim_{b \rightarrow +\infty} \exp[-e^{-s(b-n)}] - \lim_{a \rightarrow -\infty} \exp[-e^{-s(a-n)}]$$
$$= 1 + 0 = 1$$

So

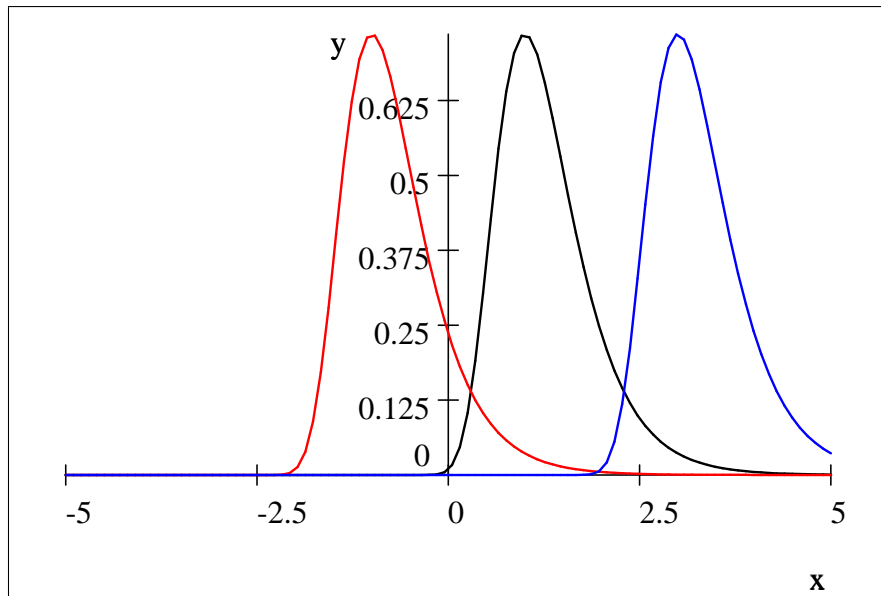
$$f(x) = se^{-s(x-n)} \exp[-e^{-s(x-n)}]$$

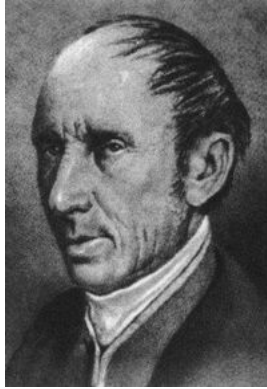
is a density function. It is called the Extreme Value Distribution. It is the foundation of logit models of choice.

Here are three examples of the Extreme Value Distribution. All have $n = 0$, the blue is $s = 2$, the red is $s = .5$ and the black is $s = 1$.



Here are three more examples. All have $s = 2$, the red is $n = -1$, the blue is $n = 3$ and the black is $n = 1$.





Example 4

$$\begin{aligned}
 f(x) &= \left(\frac{1}{\pi}\right) \frac{1}{1+(x-u)^2} \quad -\infty < x < \infty \\
 f(x) &\geq 0 \quad \forall x \\
 \int_{-\infty}^{\infty} f(x)dx &= \left(\frac{1}{\pi}\right) \int_{-\infty}^{\infty} \frac{1}{1+(x-u)^2} dx \\
 &= \left(\frac{1}{\pi}\right) \lim_{b \rightarrow \infty} \arctan(b-u) - \left(\frac{1}{\pi}\right) \lim_{a \rightarrow -\infty} \arctan(a-u)
 \end{aligned}$$

Note $\arctan(\cdot) = \tan^{-1}(\cdot)$.

$$\int_{-\infty}^{\infty} f(x)dx = \left(\frac{1}{\pi}\right) \left(\frac{\pi}{2}\right) - \left(\frac{1}{\pi}\right) \left(-\frac{\pi}{2}\right) = \frac{1}{2} + \frac{1}{2} = 1$$

$f(x) = \left(\frac{1}{\pi}\right) \frac{1}{1+(x-u)^2}$ is called the Cauchy distribution.²

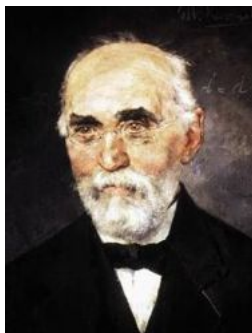
Cauchy-Lourentz Distribution. from http://en.wikipedia.org/wiki/Augustin_Cauchy,
 October 2, 2007

Augustin Louis Cauchy

Notice the great comb-over.

²This is actually a restricted Cauchy. Incorporating a scale parameter $\beta > 0$ the *Cauchy probability density function* is

$$f(x) = \frac{1}{\pi\beta \left(1 + \left(\frac{x-\mu}{\beta}\right)^2\right)}$$



Born 21 August 1789, Dijon, France

Died, 23 May 1857 (aged 67)

Now you might ask who Lorentz is, http://en.wikipedia.org/wiki/Hendrik_Lorentz,
October 2, 2007

Hendrik Antoon Lorentz

Born July 18, 1853, Arnhem, Netherlands

Died February 4, 1928 (aged 74)

Nobel Prize for Physics (1902)

The median of this distribution is μ . The Cauchy probability density function is symmetric about μ and has a unique maximum at u .

If $u = 0$ it simplifies to the Willy/Marshall distribution. (Willy and Marshall were two former students.)

$$f(x) = \left(\frac{1}{\pi}\right) \frac{1}{1+x^2}$$

The Cauchy distribution is kind of cool: it does not have a mean or a variance.³

1.2

³Or, for that matter, any *moments*. We have yet to talk about moments.

1.3 Given the density function, $f(x)$,

1.4

1.5

$$\Pr(a \leq x \leq b) = \int_a^b f(x) dx$$

For example 1

$$\begin{aligned} f(x) &= 0.222x \quad 0 \leq x \leq 3 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

so

$$\begin{aligned} \Pr(1 \leq x \leq 2) &= \int_1^2 (0.222x) dx = 0.111x^2 \Big|_1^2 \\ &= 0.111 [4 - 1] = 0.333 = \frac{1}{3} \end{aligned}$$

For example 2

$$\begin{aligned} f(x) &= \left(\frac{1}{18}\right)(3 + 2x) \quad 2 \leq x \leq 4 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

so

$$\begin{aligned} \Pr(2 < x < 3) &= \Pr(2 \leq x \leq 3) = \int_2^3 f(x) dx \\ &= \frac{1}{18} \int_2^3 (3 + 2x) dx = \frac{1}{18} (3x + 2x^2) \Big|_2^3 \\ &= \frac{1}{18} [(9 + 9) - (6 + 4)] \\ &= \frac{1}{18} 8 = \frac{8}{18} \end{aligned}$$

For example 3

$$\begin{aligned} f(x) &= se^{-s(x-n)} \exp \left[-e^{-s(x-n)} \right] \\ s &> 0 \end{aligned}$$

so

$$\begin{aligned} \Pr(-\infty < x < b) &= \Pr(x \leq b) \\ &= \int_{-\infty}^b f(x) dx = \int_{-\infty}^b se^{-s(x-n)} \exp \left[-e^{-s(x-n)} \right] dx \\ &= \exp \left[-e^{-s(b-n)} \right] - \lim_{a \rightarrow -\infty} \exp \left[-e^{-s(a-n)} \right] \end{aligned}$$

but

$$\lim_{a \rightarrow -\infty} \exp \left[-e^{-s(a-n)} \right] = 0$$

because as $a \rightarrow -\infty$,

$$-s(a-n) \rightarrow \infty$$

Therefore, as $a \rightarrow -\infty$,

$$e^{-s(a-n)} \rightarrow \infty$$

So, as $a \rightarrow -\infty$,

$$-e^{-s(a-n)} \rightarrow -\infty$$

and

$$\exp \left[-e^{-s(a-n)} \right] \rightarrow 0$$

Therefore, for the Extreme Value Distribution

$$\Pr(x \leq b) = \exp \left[-e^{-s(b-n)} \right]$$

For example,

$$\begin{aligned}\Pr(x \leq n) &= \exp\left[-e^{-s(n-n)}\right] \\ &= \exp\left[-e^0\right] = \exp\left[-1\right] = e^{-1} = \frac{1}{e}\end{aligned}$$

Consider $n + \frac{.3665}{s}$ (where did this come from?) What is the probability that $x \leq n + \frac{.3665}{s}$?

$$\begin{aligned}\Pr ob\left(x \leq n + \frac{.3665}{s}\right) &= \exp\left[-e^{-s\left(n + \frac{.3665}{s} - n\right)}\right] \\ &= \exp\left[-e^{-s\left(\frac{.3665}{s}\right)}\right] = \exp\left[-e^{-.3665}\right] = .5\end{aligned}$$

What is $n + \frac{.3665}{s}$? It is the median of the EV distribution. How did I figure out that this is the median of the EV distribution? Since

$$\Pr(x \leq b) = \exp\left[-e^{-s(b-n)}\right]$$

at median

$$\Pr(x \leq median) = \exp\left[-e^{-s(median-n)}\right] = .5$$

by definition of the median. Simplify by letting $\alpha = s(median - n)$

$$\rightarrow \exp\left[-e^{-\alpha}\right] = 0.5$$

Solve for α by taking the \ln of each side

$$\rightarrow \alpha = .366512921$$

but $\alpha = s(median - n)$. Solve for

$$median = n + \frac{.366512921}{s}$$

For example 4

$$f(x) = \frac{\left(\frac{1}{\pi}\right)}{(1+x^2)}$$
$$u = 0$$

so

$$\begin{aligned}\Pr(x > 0) &= \int_0^{\infty} f(x) dx = \left(\frac{1}{\pi}\right) \int_0^{\infty} \frac{1}{1+x^2} dx \\ &= \left(\frac{1}{\pi}\right) \left[\lim_{b \rightarrow \infty} \arctan b \right] - \left(\frac{1}{\pi}\right) [\arctan 0] \\ &= \left(\frac{1}{\pi}\right) \left(\frac{\pi}{2}\right) - \left(\frac{1}{\pi}\right) 0 \\ &= \frac{1}{2} - 0 = \frac{1}{2}\end{aligned}$$

In words, zero is the median of the Willy/Marshall distribution.

1.6

1.7 Given the density function $f(x)$, what is the probability that X is less than or equal to x , where x is a specific value of X ?

Denote this probability

$$\Pr(X \leq x) = F(x) \quad -\infty < x < \infty$$

where

$$\Pr(X \leq x) = F(x) = \int_{-\infty}^x f(t) dt$$

So the probability that $x \leq b$ is

$$\Pr(X \leq b) = F(b) = \int_{-\infty}^b f(t) dt = \int_{-\infty}^b f(x) dx$$

$F(x)$ is called the cumulative density function for x .

We have already calculated the CDF (cumulative density function) for the Extreme Value Distribution and determined that

$$F(x) = \exp \left[-e^{-s(x-n)} \right]$$

What is the CDF for?

For Example 1

$$\begin{aligned} f(x) &= 0.222x \quad 0 \leq x \leq 3 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

so

$$F(x) = \int_{-\infty}^x f(t) dt$$

$$F(x) = 0 \quad \text{if } x \leq 0$$

$$F(x) = \int_0^x 0.222t \, dt = .111t^2 \Big|_0^x = .111x^2 \quad \text{if } 0 < x < 3$$

$$F(x) = 1 \quad \text{if } x \geq 3$$

For Example 2

$$\begin{aligned} f(x) &= \left(\frac{1}{18}\right)(3+2x) \quad 2 \leq x \leq 4 \\ &= 0 \text{ otherwise} \end{aligned}$$

so

$$\begin{aligned} F(x) &= 0 \text{ if } x < 2 \\ F(x) &= 1 \text{ if } x \geq 4 \end{aligned}$$

$$\begin{aligned} F(x) &= \int_2^x \left(\frac{1}{18}\right)(3+2t) dt = \left(\frac{1}{18}\right)(3t+2t^2) \Big|_2^x \\ &= \left(\frac{1}{18}\right)[(3x+2x^2) - (6+4)] \\ &= \left(\frac{1}{18}\right)[3x+2x^2-10] \text{ if } 2 \leq x \leq 4 \end{aligned}$$

For Example 3

We have already determined that for the Extreme Value Distribution

$$F(x) = \exp \left[-e^{-s(x-n)} \right]$$

For Example 4
Trig



Looking it up: The CDF for the *Cauchy* is

$$F(x) = \frac{1}{\pi} \arctan\left(\frac{x - \mu}{\beta}\right) + \frac{1}{2}$$

For the normal distribution

There is also no closed-form solution for $F(x)$ for the normal distribution. That is

$$\int_{-\infty}^x f(t)dt$$

does not have a closed-form solution if

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-u)^2/2\sigma^2}$$

However, given specific values for u and σ^2 , one can numerically solve $\int_{-\infty}^x f(t)dt$ for any x . Most statistics books provide tables for x given $u = 0$ and $\sigma^2 = 1$ (the standard normal). Now there are web sites that provide these tables interactively, you plug in u, σ^2 and x and out comes the probability that $X \leq x$.

1.8

1.9 Note that if

$$F(a) = \Pr(x < a)$$

then

$$1 - F(a) = \Pr(x > a)$$

because

$$F(a) + (1 - F(a)) = 1$$

1.10 Measures of central tendency

Measures of central tendency are ways to describe one aspect of a distribution, $f(x)$. Three measures of central tendency are:

- mean (expected value)
- median
- mode

The mean (expected value) of any continuous random variable x with distribution $f(x)$ is defined as

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

What does $E(X)$ mean (no pun intended)? If one randomly sampled one X , one would not expect it to be $E(X)$, so it is not the number you "expect." But, if one randomly sampled N X 's, one would expect the average value of the sampled x 's to $\rightarrow E(X)$ as $N \rightarrow \infty$.

The median of a continuous random variable x with density function $f(x)$ is defined as u where

$$\int_{-\infty}^u f(x) dx = \frac{1}{2} = \int_u^{\infty} f(x) dx$$

Given my definition of the median, Is the median always a unique number, or can it be a range? Should we change my definition so that it only exists when μ is unique?

If a density function has a unique global max, the value of x that $\max f(x)$ is called the mode. Very loosely speaking, the mode is the most common value for x (remember that if x is continuously distributed the probability of observing any specific value of x is zero).

If the distribution is symmetric, mean = median.

For some distributions, mean = median = mode: e.g. the normal and the logistic.

Not that some distributions do not have a mean.

The mean of (example 1)

$$\begin{aligned} f(x) &= 0.222x \quad 0 \leq x \leq 3 \\ &= 0 \text{ otherwise} \end{aligned}$$

so

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^0 xf(x)dx + \int_0^3 xf(x)dx + \int_3^{\infty} xf(x)dx \\ &= \int_{-\infty}^0 x0dx + \int_0^3 x(.222)xdx + \int_3^{\infty} x0dx \\ &= 0 + \int_0^3 (.222)x^2dx + 0 = \frac{.222}{3}x^3|_0^3 = 0.74x^3|_0^3 \\ &= 2 \end{aligned}$$

The mean of (example 2)

$$\begin{aligned} f(x) &= \left(\frac{1}{18}\right)(3+2x) \quad 2 \leq x \leq 4 \\ &= 0 \text{ otherwise} \end{aligned}$$

is

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^2 xf(x)dx + \int_2^4 xf(x)dx + \int_4^{\infty} xf(x)dx \\ &= \int_{-\infty}^2 x0dx + \int_2^4 x\left(\frac{1}{18}\right)(3+2x)dx + \int_4^{\infty} x0dx \\ &= 0 + \left(\frac{1}{18}\right) \int_2^4 (3x+2x^2)dx + 0 \\ &= \left(\frac{1}{18}\right) \left(\frac{3}{2}x^2 + \frac{2}{3}x^3\right) \Big|_2^4 = \left(\frac{1}{18}\right) \left[\frac{3}{2}(16) + \frac{2}{3}(64) - \left(\frac{3}{2}(4) + \frac{2}{3}(8)\right)\right] \\ &= \frac{1}{18} [(24 + 42.666) - (6 + 5.333)] = \frac{1}{18} [66.666 - 11.333] \\ &= 3.074 \end{aligned}$$

For the Extreme Value Distribution (example 3)

$$f(x) = se^{-s(x-n)} \exp \left[-e^{-s(x-n)} \right]$$

so

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} xse^{-s(x-n)} \exp \left[-e^{-s(x-n)} \right] dx \\ &= \int_{-\infty}^{\infty} xm'(x)e^{m(x)}dx =? \end{aligned}$$

where

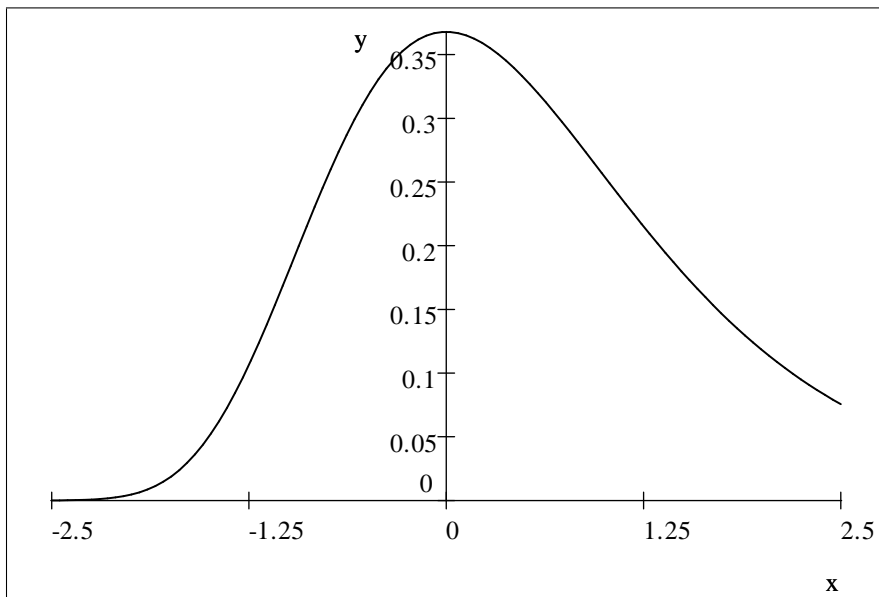
$$m(x) = -e^{-s(x-n)}$$

I know that $E(X) = n + \frac{\gamma}{s}$ where γ is the Euler constant (0.577). But I have been unable to derive it analytically. Can you?

Can you do it for a specific values of n and s ? For example, if $n = 0$ and $s = 1$,

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} xe^{-x} \exp[-e^{-x}] dx$$

Given $E(X) = n + \frac{\gamma}{s}$, if $s = 1$ and $n = 0$, $E(X) = n + \frac{\gamma}{s} = \gamma = 0.577\dots$. Looking at the graph of $e^{-x} \exp[-e^{-x}]$, one might guesstimate that $E(X) = 0.577\dots$, is a bit greater than zero.



In this special case ($n = 0$ and $s = 1$), $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} xe^{-x} \exp[-e^{-x}] dx,$

my math software can figure out the correct answer. Note that below I am taking the definite integral over an every increasing range, and that the answer is approaching γ . (My software will get mad if I plug in positive and negative infinity)

$$\int_{-1}^1 xe^{-x} \exp[-e^{-x}] dx = 1.7505 \times 10^{-2}$$

$$\int_{-2}^2 xe^{-x} \exp[-e^{-x}] dx = 0.19448$$

$$\int_{-5}^5 xe^{-x} \exp[-e^{-x}] dx = 0.53691$$

$$\int_{-10}^{10} xe^{-x} \exp[-e^{-x}] dx = 0.57672$$

$$\int_{-20}^{20} xe^{-x} \exp[-e^{-x}] dx = 0.57722$$

If I wanted to generalize I would continue to assume $s = 1$ and try to take the definite integral for different values of n to see how this affects the answer. Hopefully, I would find that $E(X) = n + \gamma$, and, then if that works, I would start playing with different values for s .

As an aside, note that the mode is n .

For the Cauchy Distribution (example 4)

It does not have a mean. Wow. For a bit of explanation as to why, see the asides at the end of this lecture.

1.10.1 Medians

We have already determined the median for the Extreme Value Distribution (example 3) and the Willy/Marshall distribution (example 4).

If u is the median

$$\Pr(x \leq u) = .5$$

For the Extreme Value Distribution

$$f(x) = se^{-s(x-n)} \exp \left[-e^{-s(x-n)} \right]$$

Earlier we determined that

$$median = n + \frac{3665}{s}$$

Recollect that

$$mean = E(X) = n + \frac{\gamma}{s}$$

where γ is the Euler constant $\sim .577216$.

So, $mean \neq median$ and

$$mean - median = \left(n + \frac{.577216}{s} \right) - \left(n + \frac{3665}{s} \right) = \frac{.21070}{s}$$

Earlier we determined that the median for the Willy/Marshall distribution is 0. More generally the median of the Cauchy distribution is μ - it has a median, just not a mean.

Determine the mode for the four example distributions?

1.11 Measures of dispersion

One question we often ask about distributions is: How dispersed or spread out are the values of X ?

The most common measure of spread is **variance**. The variance is a measure of dispersion around the mean ($E(X)$).

$$\begin{aligned} \text{Variance} &\equiv \sigma^2 \equiv E \left[(x - E(X))^2 \right] \\ &\equiv \text{expected value of } (x - E(X))^2 \end{aligned}$$

Another possible measure of the dispersion around the mean is

$$E [|x - E(X)|]$$

One could alternatively define measures of dispersion around the median or mode.

It can be shown that if $f(x)$ is the density function for x and $g(x)$ is some function of x

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

(this is a formula we will use often.)⁴ We used a special case of this formula to get $E(X) = \text{mean}$. That is, if $g(x) = x$

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

We can also use it to get

$$\sigma^2 \equiv E[(x - E(X))^2]$$

In this case, $g(x) = (x - E(X))^2$ and

$$\sigma^2 \equiv E[(x - E(X))^2] = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx$$

⁴Investigate why it is true.

Let's calculate the variance of x for our earlier examples.

For example 1

$$\begin{aligned} f(x) &= 0.222x \quad 0 \leq x \leq 3 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

The variance is

$$\sigma^2 = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

Since $E(X) = 2$,

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - 2)^2 f(x) dx \\ &= \int_{-\infty}^0 (x - 2)^2 0 dx + \int_0^3 (x - 2)^2 (.222x) dx + \int_3^{\infty} (x - 2)^2 0 dx \\ &= \int_0^3 (x - 2)^2 (.222x) dx \\ &= .5 \end{aligned}$$

So, what does this mean? In words, what does a variance of .5 mean?

For example 2

$$\begin{aligned} f(x) &= \left(\frac{1}{18}\right)(3+2x) \quad 2 \leq x \leq 4 \\ &= 0 \text{ otherwise} \end{aligned}$$

Since $E(X) = 3.074$, the variance of x is

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx \\ &= \int_{-\infty}^{\infty} (x - 3.074)^2 f(x) dx \\ &= \int_{-\infty}^2 (x - 3.074)^2 0 dx + \int_2^4 (x - 3.074)^2 \left(\frac{1}{18}\right) (3 + 2x) dx + \int_4^{\infty} (x - 3.074)^2 0 dx \\ &= \int_2^4 (x - 3.074)^2 \left(\frac{1}{18}\right) (3 + 2x) dx \\ &= .327846 \end{aligned}$$

For Example 3

Can we figure out the variance for the Extreme Value Distribution?

$$f(x) = se^{-s(x-n)} \exp \left[-e^{-s(x-n)} \right]$$

Since $E(X) = n + \frac{\gamma}{s}$, where γ is Euler's constant

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - E(X))^2 se^{-s(x-n)} \exp \left[-e^{-s(x-n)} \right] dx \\ &= \int_{-\infty}^{\infty} \left(x - \left(n + \frac{\gamma}{s} \right) \right)^2 se^{-s(x-n)} \exp \left[-e^{-s(x-n)} \right] dx \end{aligned}$$

Luckily for us, someone (?) has already determined that

$$\sigma^2 = \frac{\pi^2}{6s^2}$$

Example 4 (the Cauchy distribution). Since it does not have an $E[X]$, it cannot have a well-defined variance because the variance is a function of the $E[X]$.

For the normal density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

the variance is simply σ^2 .

1.12 Some asides

1.12.1 Is there a way to derive the $E[X]$ directly from the CDF?

The answer is yes (from MGB), as there must be:

$$E[X] = \int_0^{+\infty} [1 - F(x)] dx - \int_{-\infty}^0 F(x) dx$$

Can you intuite this? Try it out and see if it works. If I remember correctly, this formula is useful if the one has a mixed distribution (discrete and continuous).

Maybe, you could use it to get $E[X]$ for the Extreme Value distribution, where $F(x) = \exp[-e^{-s(x-n)}]$, $E[X] = \int_0^{+\infty} [1 - F(x)] dx - \int_{-\infty}^0 F(x) dx = \int_0^{+\infty} [1 - \exp[-e^{-s(x-n)}]] dx - \int_{-\infty}^0 \exp[-e^{-s(x-n)}] dx$. NOPE

1.12.2 Some probabilities of interest

For some distributions, one you determine $\Pr[(E[X] - \sigma) \leq x \leq (E[X] + \sigma)]$ where σ is defined as the variance, or $\Pr[(E[X] - 2\sigma) \leq x \leq (E[X] + 2\sigma)]$. The first is the probability that a randomly drawn X lies between $(E[X] - \sigma)$ and $(E[X] + \sigma)$

For a given $f(x)$, this is the probability that

At this point, you should have the tools to figure these out? Will the value of $\Pr[(E[X] - \sigma) \leq x \leq (E[X] + \sigma)]$ be a function of $f(x)$? or be invariant to the form of the density function?

For example, in terms of the CDF, $F(X)$

$$\begin{aligned}\Pr[(E[X] - \sigma) \leq x \leq (E[X] + \sigma)] \\ = F(E[X] + \sigma) - F((E[X] - \sigma))\end{aligned}$$

It looks like it is a function of $F(X)$, so not invariant to the form of the distribution.⁵

1.12.3 $E[X]$ and variance as areas

Recollect that if $f(x)$ is a density function then $\int_{-\infty}^{+\infty} f(x)dx = 1$, where $\int_{-\infty}^{+\infty} f(x)dx$ is the area under the density function.

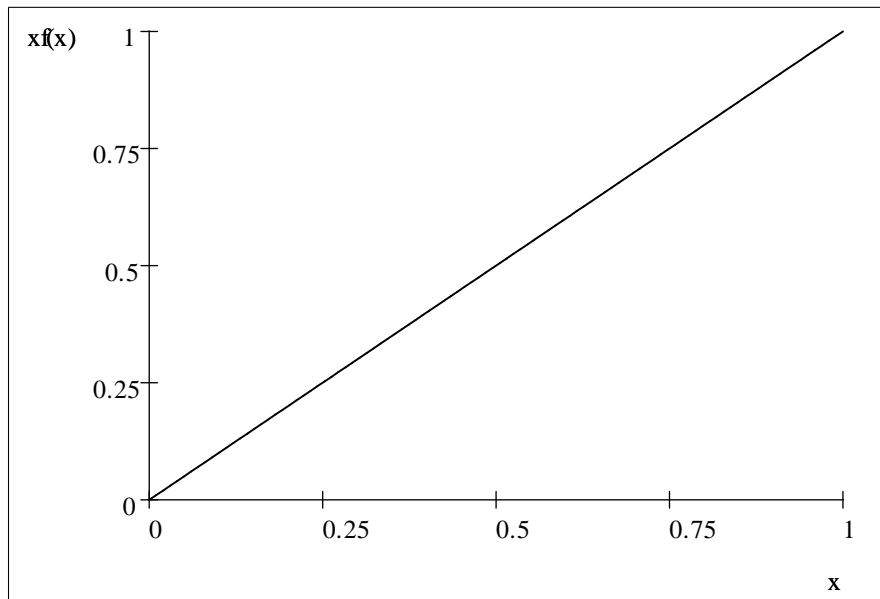
$E(X) = \int_{-\infty}^{\infty} xf(x)dx$ can be thought of as an area, the area under the function $xf(x)$.

And, variance is the area under $(x - E[X])^2f(x)$. That is, mean and variance can be viewed as areas.

So the mean will not be well defined unless the area under $xf(x)$ is well defined, and the variance will not be well defined unless the area under $(x - E[X])^2f(x)$ is well defined - this will sometimes be an issue, mostly with density functions where $f(x) > 0$ $-\infty < x < +\infty$.

Consider the uniform distribution on the unit interval, a very simple example. For this density, $xf(x)$ is a straight line from $(0, 0)$ to $(1, 1)$, and the area under it is $.5 = E[X]$

⁵Many students tend to think, incorrectly, that $\Pr[(E[X]-\sigma) \leq x \leq (E[X]+\sigma)]$ is invariant to the form of $F(X)$

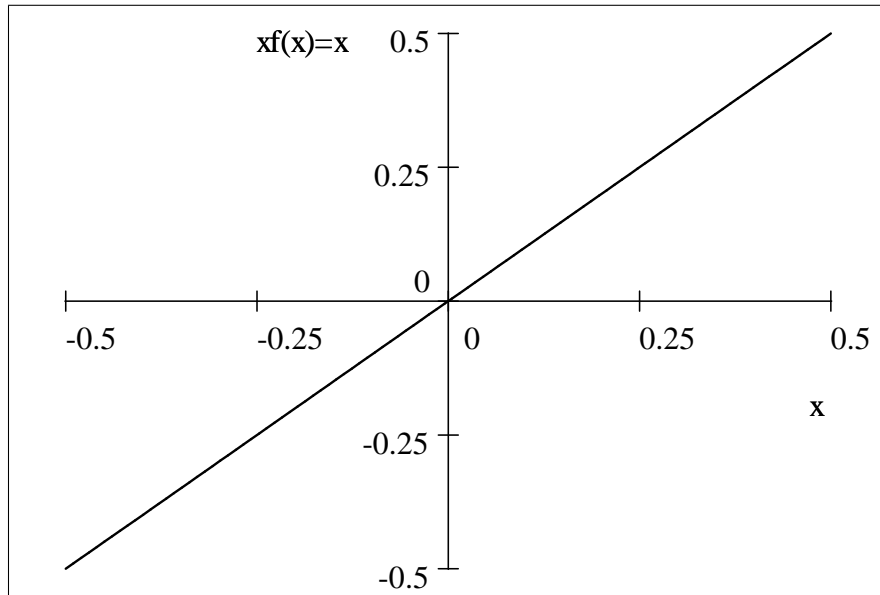


the function $xf(x) = x(1-x)$

Note that in regions where $x < 0$ and $f(x) > 0$, $xf(x) < 0$, so if the $xf(x)$ spans any of the negative region,

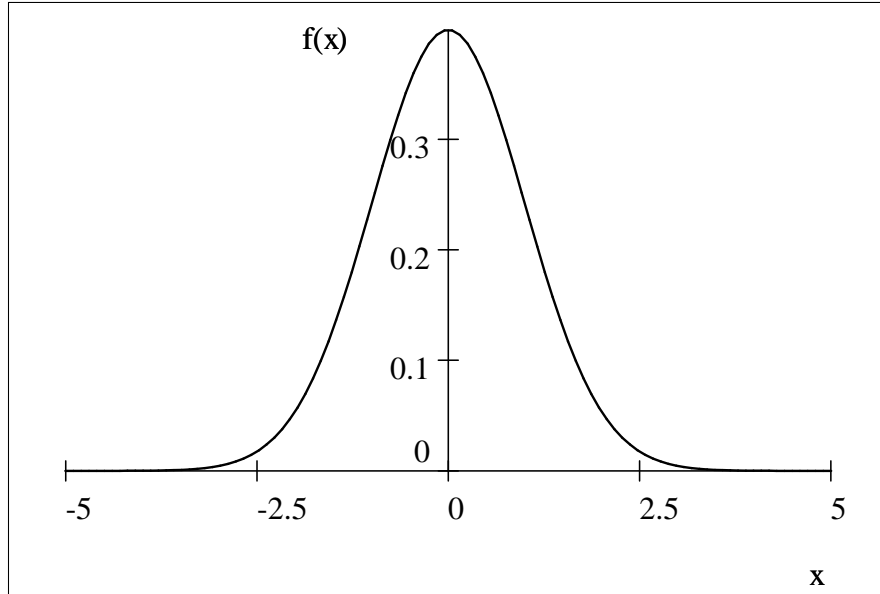
it will have negative area in those sections.

Consider, for example, the uniform distribution on the interval $(-.5, +.5)$. For this density function, $xf(x)$ is a straight line from $(-.5, -.5)$ to $(.5, .5)$; the area under it to the left of zero is $-.25$ and to the right of zero is $+.25$ and $E[X] = 0$.



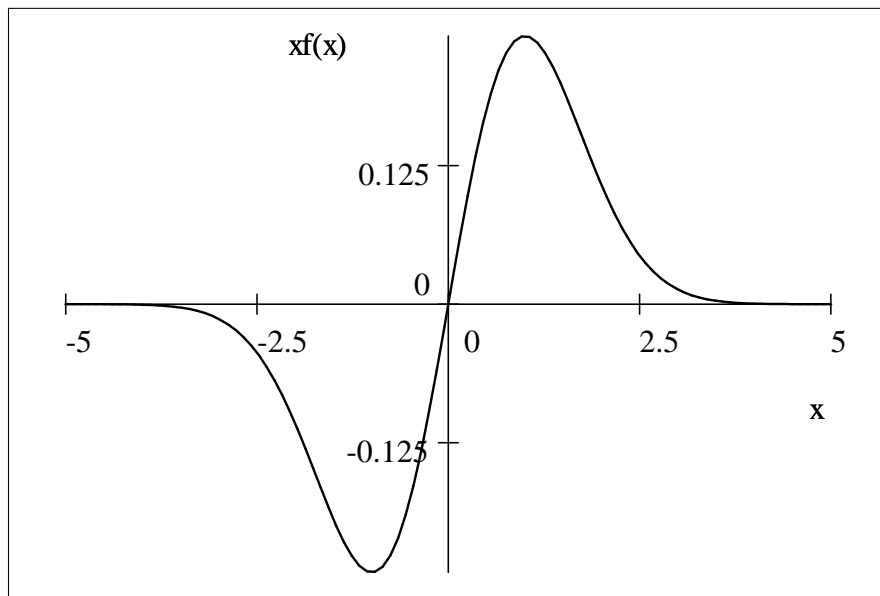
$$\int_{-.5}^{.5} x dx = 0.0$$

Consider the standard normal $f(x) = \frac{1}{\sqrt{2\pi}}e^{-(x)^2/2}$



The area under this curve is 1 - that's why it is a density function

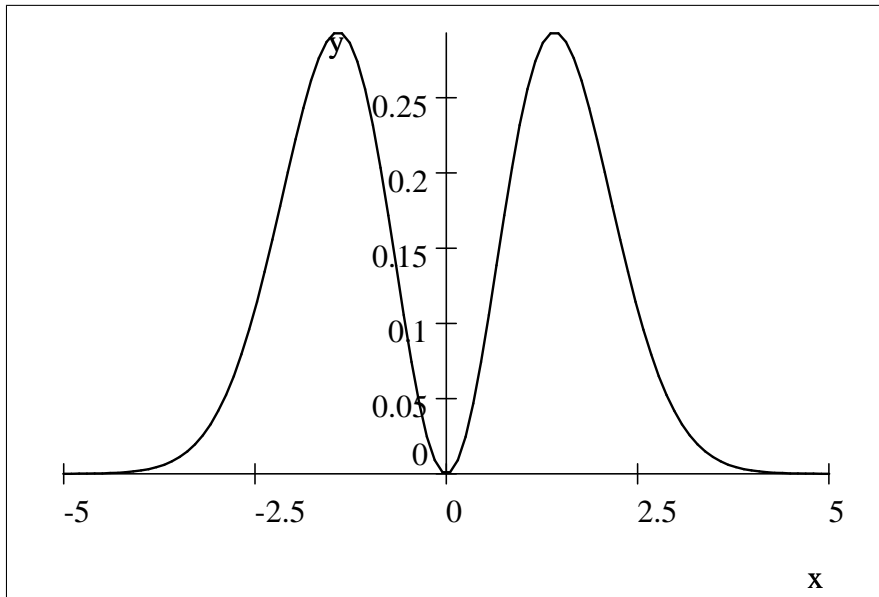
For the standard normal $xf(x) = \frac{x}{\sqrt{2\pi}}e^{-(x)^2/2}$



and the "area under this is zero, $\int_{-10}^{10} (\frac{x}{\sqrt{2\pi}}e^{-(x)^2/2})dx = 0$, so the correct answer

for the mean of the standard normal.

How about the variance (which we know is 1). It is the area under $(x - 0)^2 f(x) = \frac{x^2}{\sqrt{2\pi}} e^{-(x)^2/2}$

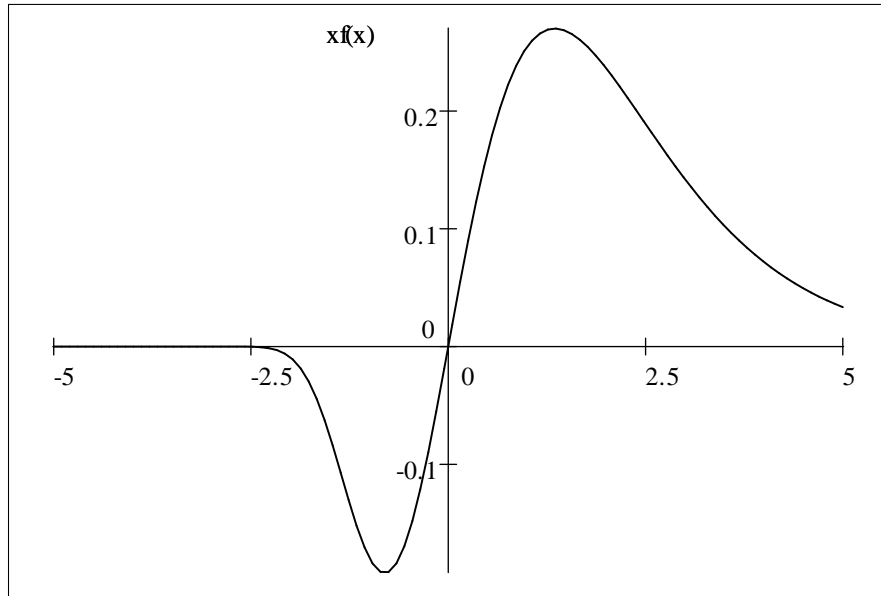


$$\int_{-5}^5 \left(\frac{x^2}{\sqrt{2\pi}} e^{-(x)^2/2} \right) dx = 0.99998, \text{ wow, the correct answer.}$$

Further note that we can always write

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{-\infty}^0 x f(x) dx + \int_0^{+\infty} x f(x) dx \\ &= \int_0^{+\infty} x f(x) dx - \int_{-\infty}^0 |x| f(x) dx \end{aligned}$$

for the extreme value ($n = 0$ and $s = 1$), $E(X)$ is the area under $xe^{-x} \exp[-e^{-x}]$



As from before $\int_{-25}^{25} x e^{-x} \exp[-e^{-x}] dx = 0.577 22$

1.12.4 Density functions where $f(x) > 0 -\infty < x < +\infty$

1.12.5

The normal, the Cauchy and the Extreme Value are all density function where every value of X has positive density.

Imagine you are trying to create a new density function where $f(x) > 0 \forall x$. The area under many functions that meet the condition $f(x) > 0 \forall x$ have $\int_{-\infty}^{+\infty} f(x) dx = +\infty$, so cannot be density functions.

One must choose carefully: one must choose a function so that as one integrates over larger and larger ranges of X , the area under the function must forever increase, but at the same time the area can never be greater than one - neat trick.

The normal, the Cauchy and the Extreme value, since they all are density functions, do the trick. But the fact that the area under $f(x)$ is well defined, does not imply that the areas under $xf(x)$ and $(x - E[X])^2 f(x)$ are well defined.⁶

These areas are well defined for the normal and the Extreme value, but they are not for the Cauchy - that is why the Cauchy does not have a mean.

Put loosely, the reason the Cauchy does not have a mean is that $\int_{-\infty}^0 xf(x) dx = -\infty$ and $\int_0^{+\infty} xf(x) dx = +\infty$ - the tails of $xf(x)$ are *too fat*.

⁶Note that if the area under $xf(x)$ is not well defined, then $(x - E[X])^2 f(x)$ is not well defined.

Why the mean of the Cauchy distribution is undefined

If a [probability distribution](#) has a [density function](#) $f(x)$ then the mean or [expected value](#) is

$$\int_{-\infty}^{\infty} x f(x) dx. \quad (1)$$

The question is now whether this is the same thing as

$$\int_0^{\infty} x f(x) dx - \int_{-\infty}^0 |x| f(x) dx. \quad (2)$$

If at most one of the two terms in (2) is infinite, then (1) is the same as (2). But in the case of the Cauchy distribution, both the positive and negative terms of (2) are infinite. This means (2) is undefined. Moreover, if (1) is construed as a [Lebesgue integral](#), then (1) is also undefined, since (1) is then defined simply as the difference (2) between positive and negative parts.

However, if (1) is construed as an [improper integral](#) rather than a Lebesgue integral, then (2) is undefined, and (1) is not necessarily well-defined. We may take (1) to mean

$$\lim_{a \rightarrow \infty} \int_{-a}^a x f(x) dx,$$

and this is its [Cauchy principal value](#), which is zero, but we could also take (1) to mean, for example,

$$\lim_{a \rightarrow \infty} \int_{-2a}^a x f(x) dx,$$

which is *not* zero, as can be seen easily by computing the integral.

Various results in probability theory about [expected values](#), such as the [strong law of large numbers](#), will not work in such cases.

Also, the sample mean of a random sample taken from a Cauchy distribution is no better than a single observation, because the chances of including extreme values is high. However, the sample median, which is not affected by extreme values, can be used as a measure of central tendency.

The following is from Wikipedia.

Go figure!