An Economic/Statistical Application of Integration: Probability Theory

What is a random variable?

x is a random variable if it has a known distribution. That is, x is a random variable if $\forall$ a and b one can determine the probability that $a \le x \le b$.

> Note that x takes specific values (e.g., if x is weight, each of us has a specific weight but weight, in the population, has some distribution.

A well known statistics book, *Introduction to the Theory of Statistics* (Mood & Graybill) defines a continuous random variable as follows.

The variable X is a one-dimensional, continuous random variable if there exists a function $f(x)$ such that $f(x) \ge 0 \; \forall \; x$ in the interval $-\infty \le x \le \infty$, and the probability that $(a \le x \le b)$ is

$$\text{Prob } (a \le x \le b) = \int_a^b f(x)\, dx .$$

The function $f(x)$ is called a "*density function*" (or a "*probability density function*").
Any function, $f(x)$, can serve as a density function as long as

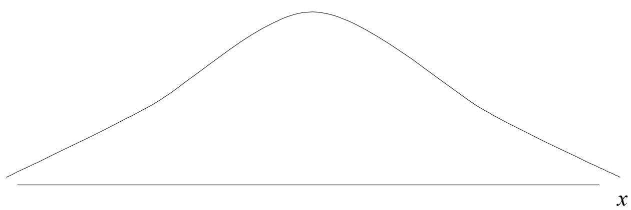
$$f(x) \ge 0, \quad -\infty \le x \le \infty$$

and

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

Examples of Density Functions

The normal density function



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

is a well known density function

where σ and u are parameters in the density function.

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

Lets start more simply.

Make up a simple density function.

Note that it is not necessary that $f(x) > 0 \ \forall \ x$.

Consider the Following Four Density Functions

- **Example 1**

Assume

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

and

$$g(x) = 3x \quad 0 \le x \le 3$$

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

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

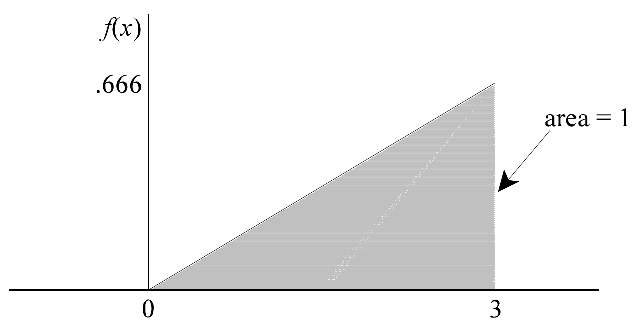
$$f(x) = \frac{3x}{13.5} \text{ is.}$$

That is

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

and

$$g(x) = .222x \quad \text{if} \quad 0 \le x \le 3 .$$



- **Example 2**

$$f(x) = \left(\frac{1}{18}\right)(3 + 2x) \quad 2 \le x \le 4$$

$$f(x) = 0 \quad \text{otherwise}$$

Is this a density function?

$$f(x) \ge 0 \;\forall\, x$$

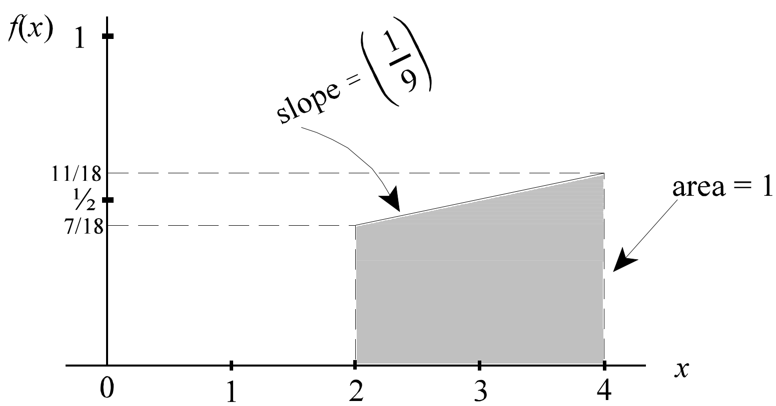
and

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{2} 0\,dx + \int_{2}^{4} \left(\frac{1}{18}\right)(3+2x)\,dx + \int_{4}^{\infty} 0\,dx$$

$$= \frac{1}{18}\int_{2}^{4}(3+2x)\,dx = \frac{1}{18}\left(3x+x^2\right)\Big|_2^4 = \frac{1}{18}\left[(12+16)-(6+4)\right].$$

$$= 1$$

So, yes.



- **Example 3**

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

$$s > 0$$

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

$$f(x) \geq 0 \;\forall\, x$$

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

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

Why? Recollect that $\int m'(x)\, e^{m(x)} dx = e^{m(x)}$

So

$$\int_{-\infty}^{\infty} f(x)\,dx = \lim_{b \to +\infty} \exp\left[-e^{-s(b-n)}\right]$$

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

$$= 1 + 0 = 1$$

So $f(x) = se^{-s(x-n)} \exp\left[-e^{-s(x-n)}\right]$ is a density function.

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

- **Example 4**

$$f(x) = \left(\frac{1}{\pi}\right) \frac{1}{1+(x-u)^2} \qquad -\infty < x < \infty$$

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

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

$$= \left(\frac{1}{\pi}\right) \lim_{b \to \infty} \arctan(b-u) - \left(\frac{1}{\pi}\right) \lim_{a \to -\infty} \arctan(a-u)$$

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

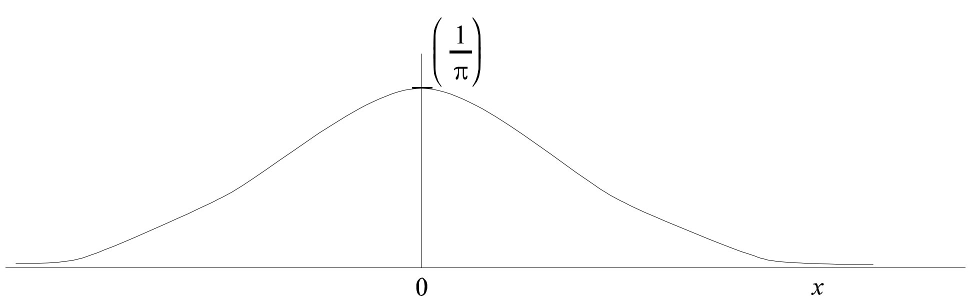
$$= \left(\frac{1}{\pi}\right)\left(\frac{\pi}{2}\right) - \left(\frac{1}{\pi}\right)\left(-\frac{\pi}{2}\right) = \frac{1}{2} + \frac{1}{2} = 1$$

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

If $u = 0$ it simplifies to the "Willy/Marshall" distribution

Willy and Marshal were two former students.

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



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

$$\text{Prob }(a \le x \le b) = \int_a^b f(x)\,dx.$$

- **For example 1**

$$f(x) = .222x \quad 0 \le x \le 3$$
$$= 0 \text{ otherwise}$$

$$\text{Prob }(1 \le x \le 2) = \int_1^2 (.222x)\,dx = .111x^2\Big|_1^2$$

$$= .111\big[(4) - 1\big] = .333 = \frac{1}{3}.$$

- **For example 2**

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

$$\text{Prob }(2 < x < 3) = \text{Prob }(2 \le x \le 3) = \int_2^3 f(x)\,dx$$

$$= \frac{1}{18}\int_2^3 (3 + 2x)\,dx = \frac{1}{18}(3x + x^2)\Big|_2^3$$

$$= \frac{1}{18}\big[(9 + 9) - (6 + 4)\big]$$

$$= \frac{1}{18}\,8 = \frac{8}{18}.$$

- **For example 3**

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

$$s > 0$$

$$\text{Prob } (-\infty < x < b) \equiv \text{ Prob } (x \leq b)$$

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

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

but

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

because

as $a \to -\infty$, $(-s(a-n)) \to \infty$

therefore

as $a \to -\infty$, $\left[e^{-s(a-n)}\right] \to \infty$

so, as $a \to -\infty$

$\left[-e^{-s(a-n)}\right] \to -\infty$

and

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

Therefore, for the Extreme Value Distribution

$$\text{Prob}\ (x \le b) = \exp\left[-e^{-s(b-n)}\right].$$

For example

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

$$= \exp\left[-e^{0}\right] = \exp\left[-1\right] = e^{-1} = \frac{1}{e}$$

What is the probability that $x \le n + .3665/s$?

$$\text{Prob}\ (x \le n + .3665/s) = \exp\left[-e^{-s(n+(3.665/s)-n)}\right] = \exp\left[-e^{\left[-s(.3665/s)\right]}\right]$$

$$= \exp\left[-e^{-.3665}\right] = .5$$

What is $(n + .3665/s)$?
It is the median of the EV distribution.

How did I figure out that the median of the EV distribution is $(n + .3665/s)$?

Since

$$\text{Prob}\ (x \le b) = \exp\left[-e^{-s(b-n)}\right]$$

at median

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

by definition of the median.

Simplify by letting $\alpha = s(\text{median} - n)$

$$\Rightarrow \exp\left[-e^{-\alpha}\right] = .5$$

solve for α by taking the ln of each side

$$\Rightarrow \alpha = .366512921$$

but $\alpha = s(\text{median} - n)$

solve for

$$\text{median} = n + .366512921/s.$$

- **For example 4**

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

$$u = 0$$

$$\text{Prob } (x > 0) = \int_0^\infty f(x)\,dx = \left(\frac{1}{\pi}\right) \int_0^\infty \frac{1}{1+x^2}\,dx$$

$$= \left(\frac{1}{\pi}\right) \left[\lim_{b \to \infty} \arctan b\right] - \left(\frac{1}{\pi}\right) [\arctan 0]$$

$$= \left(\frac{1}{\pi}\right) \left(\frac{\pi}{2}\right) - \left(\frac{1}{\pi}\right) 0$$

$$= \frac{1}{2} - 0 = \frac{1}{2}.$$

In words, zero is the median of the distribution.

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

$$\text{Prob}\ (X \le x) \equiv F(x) \qquad -\infty < x < \infty$$

$$\text{Prob}\ (X \le x) \equiv F(x) = \int_{-\infty}^{x} f(t)\, dt$$

So the probability that $x \le b$ is

$$\text{Prob}\ (X \le b) = F(b) = \int_{-\infty}^{b} f(t)\, dt = \int_{-\infty}^{b} f(x)\, dx$$

$F(x)$ 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?

- **Example 1**

$$f(x) = .222x \quad 0 \le x \le 3$$
$$= 0 \quad \text{otherwise}$$

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

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

$$F(x) = \int_{0}^{x} .222t\, dt = .111t^2 \Big|_0^x = .111x^2 \quad \text{if} \quad 0 < x < 3$$

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

- **Example 2**

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

$$F(x) = 0 \quad \text{if} \quad x < 2$$

$$F(x) = 1 \quad \text{if} \quad x \ge 4$$

$$F(x) = \int_{2}^{x} \left(\frac{1}{18}\right)(3 + 2t)\, dt = \left(\frac{1}{18}\right)\left(3t + t^2\right)\Big|_{2}^{x}$$

$$= \left(\frac{1}{18}\right)\left[\left(3x + x^2\right) - (6 + 4)\right]$$

$$= \left(\frac{1}{18}\right)\left(3x + x^2 - 10\right) \quad \text{if} \quad 2 \le x \le 4.$$

- **Example 3**

We have already determined that for the Extreme Value Distribution

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

- **Example 4**

Trig ☹

There is not a 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

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

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).

> Note that if
>
> $$F(a) = \text{Prob}\,(x < a)$$
>
> then
>
> $$1 - F(a) = \text{Prob}\,(x > a)$$
>
> because
>
> $$F(a) + (1 - F(a)) = 1\,.$$

Now consider 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, and
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? If one randomly sampled one x, one would not expect it to be $E(x)$. But, if one randomly sampled N x's, one would expect the average value of the sampled x's to $\to E(x)$ as $N \to \infty$.

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

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

If a density function has a unique global max, the value of x that max $f(x)$ is called the mode. Loosely speaking, the mode is the most common value for x (remember that since 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.

- **The mean of (example 1)**

$$f(x) = .222x \quad 0 \le x \le 3$$
$$= 0 \ \text{otherwise}$$

is

$$E(x) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-\infty}^{0} xf(x)\,dx + \int_{0}^{3} xf(x)\,dx + \int_{3}^{\infty} xf(x)\,dx$$

$$= \int_{-\infty}^{0} x0\,dx + \int_{0}^{3} x(.222)x dx + \int_{3}^{\infty} x0\,dx$$

$$= 0 + \int_{0}^{3} .222x^2 dx + 0 = \frac{.222}{3}x^3\Big|_0^3 = .074\,x^3\Big|_0^3$$

$$= 2$$

- **The mean of (example 2)**

$$f(x) = \left(\frac{1}{18}\right)(3+2x) \quad 2 \le x \le 4$$

$$f(x) = 0 \quad \text{otherwise}$$

is

$$E(x) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-\infty}^{2} xf(x)\,dx + \int_{2}^{4} xf(x)\,dx + \int_{4}^{\infty} xf(x)\,dx$$

$$= \int_{-\infty}^{2} x0\,dx + \int_{2}^{4} x\left(\frac{1}{18}\right)(3+2x)\,dx + \int_{4}^{\infty} x0\,dx$$

$$= 0 + \left(\frac{1}{18}\right)\int_{2}^{4} \left(3x+2x^2\right)dx + 0$$

$$= \left(\frac{1}{18}\right)\left(\frac{3}{2}x^2 + \frac{2}{3}x^3\right)\Bigg|_2^4 = \frac{1}{18}\left[\frac{3}{2}(16) + \frac{2}{3}(64) - \left(\frac{3}{2}(4) + \frac{2}{3}(8)\right)\right]$$

$$= \frac{1}{18}\left[(24+42.666) - (6+5.333)\right] = \frac{1}{18}\left[66.666 - 11.333\right]$$

$$= 3.074.$$

- **For the Extreme Value Distribution (example 3)**

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

and

$$E(x) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-\infty}^{\infty} xse^{-s(x-n)} \exp\left[-e^{-s(x-n)}\right] dx$$

$$= \int_{-\infty}^{\infty} xm'(x)e^{m(x)}\,dx$$

where

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

I know that $E(x) = n + \gamma/s$ where γ is the Euler constant (2.577). But I have been unable to derive it analytically. Can you?

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

$$\text{Prob}\,(x \le u) = .5$$

For the Extreme Value Distribution

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

on page 9 we determined that

$$\text{median} = n + .3665/s$$

recollect that

$$\text{mean} = E(x) = n + \delta/s \text{ where } \delta \text{ is the Euler constant } \sim .577216.$$

So, mean ≠ median

and $\text{mean} - \text{median} = \left(n + \dfrac{.577216}{s}\right) - \left(n + \dfrac{0.3665}{s}\right) = \dfrac{.21070}{s}.$

On page 10 we determined that the median for the Willy/Marshall distribution is 0.

Determine the mode for the first three example distributions?

One question we often ask about distributions is: How <u>dispersed</u> or <u>spread out</u> are the values of x?

The most common measure of spread is <u>variance</u>.

The Variance is a measure of dispersion around the mean $(E(x))$.

Variance $\equiv \sigma^2 = E\left[(x - E(x))^2\right]$

$\equiv$ expected value of $(x - E(x))^2$.

> Another possible measure of the dispersion around the mean is $E\left[\,|x - E(x)|\,\right]$.

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

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

We used a special case of this relationship to get $E[x] \equiv \text{mean}$.
That is, if $g(x) = x$

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

We can also use it to get

$$\sigma^2 = E\left[(x - E(x))^2\right]$$

in this case $g(x) = (x - E(x))^2$

and

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

The variance of x if

$$f(x) = .222x \quad 0 \le x \le 3 \qquad \text{(example 1)}$$
$$f(x) = 0 \qquad \text{otherwise}$$

is

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

since $E(x) = 2$ (page 15)

$$= \int_{-\infty}^{\infty} (x-2)^2 f(x)\, dx$$

$$= \int_{-\infty}^{0} (x-2)^2\, 0\, dx$$

$$+ \int_{0}^{3} (x-2)^2 (.222x)\, dx$$

$$+ \int_{3}^{\infty} (x-2)^2\, 0\, dx$$

$$= \int_{0}^{3} (x-2)^2 (.222x)\, dx$$

$$= .5$$

The variance of *x* if

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

is

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

since $E(x) = 3.074$ (page 15)

$$= \int_{-\infty}^{\infty} (x - 3.074)^2 f(x)\, dx$$

$$= \int_{-\infty}^{2} (x - 3.074)^2\, 0\, dx + \int_{2}^{4} (x - 3.074)^2 \left(\frac{1}{18}\right)(3+2x)\, dx$$

$$+ \int_{4}^{\infty} (x - 3.074)^2\, 0\, dx$$

$$= \int_{2}^{4} (x - 3.074)^2 \left(\frac{1}{18}\right)(3+2x)\, dx$$

$$= .327846$$

Can we figure out the variance for the Extreme Value Distribution
$f(x) = se^{-s(x-n)} \exp\left[-e^{-s(x-n)}\right]$?

$$\sigma^2 = \int_{-\infty}^{\infty} (x - E(x))^2 \, se^{-s(x-n)} \exp\left[-e^{-s(x-n)}\right] dx$$

since $E(x) = n + \delta / s$
where δ is the Euler constant

$$= \int_{-\infty}^{\infty} (x - (n + \delta/s))^2 \, se^{-s(x-n)} \exp\left[-e^{-s(x-n)}\right] dx.$$

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

$$\sigma^2 = \pi^2 / 6s^2.$$

For the normal density function

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

the variance is σ^2.