Lecture 3: Random variables and their distributions

Definition 1.4.1.

A random variable X is a function from S into a subset of $\mathscr{R}$ such that for any Borel set $B \subset \mathscr{R}$ $\{X \in B\} = \{\omega \in S : X(\omega) \in B\}$ is an event.

- It is convenient to deal with a summary variable than the elements in the original sample space.
- The range of a random variables is simpler than S.
- In some problems, there is a natural random variable; e.g., the number of accidents, number of successes, etc. In other cases, we may define a random variable according to our interests.

Example 1.4.3. (Toss a coin 3 times)

outcome	hhh	hht	hth	thh	tth	tht	htt	ttt
X: number of heads	3	2	2	2	1	1	1	0

- The range of X, $\{0, 1, 2, 3\}$, is simpler than S.
- X treats hht, hth, thh the same, and tth, tht, htt the same.

Lecture 3: Random variables and their distributions

Definition 1.4.1.

A random variable X is a function from S into a subset of $\mathscr{R}$ such that for any Borel set $B \subset \mathscr{R}$ $\{X \in B\} = \{\omega \in S : X(\omega) \in B\}$ is an event.

- It is convenient to deal with a summary variable than the elements in the original sample space.
- The range of a random variables is simpler than S.
- In some problems, there is a natural random variable; e.g., the number of accidents, number of successes, etc. In other cases, we may define a random variable according to our interests.

Example 1.4.3. (Toss a coin 3 times)

outcome	hhh	hht	hth	thh	tth	tht	htt	ttt
X: number of heads	3	2	2	2	1	1	1	0

- The range of X, $\{0,1,2,3\}$, is simpler than S.
- X treats hht, hth, thh the same, and tth, tht, htt the same.

Lecture 3: Random variables and their distributions

Definition 1.4.1.

A random variable X is a function from S into a subset of $\mathscr{R}$ such that for any Borel set $B \subset \mathscr{R}$ $\{X \in B\} = \{\omega \in S : X(\omega) \in B\}$ is an event.

- It is convenient to deal with a summary variable than the elements in the original sample space.
- The range of a random variables is simpler than S.
- In some problems, there is a natural random variable; e.g., the number of accidents, number of successes, etc. In other cases, we may define a random variable according to our interests.

Example 1.4.3. (Toss a coin 3 times)

outcome	hhh	hht	hth	thh	tth	tht	htt	ttt
X: number of heads	3	2	2	2	1	1	1	0

- The range of X, $\{0,1,2,3\}$, is simpler than S.
- X treats hht, hth, thh the same, and tth, tht, htt the same.

Induced probability

The induced probability of X is

$$P_X(B) = P(X \in B) = P(\{\omega \in S : X(\omega) \in B\})$$

The probability P_X is called the distribution of X.

Example 1.4.3 (continued)

If the coin is fair, then

x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

A table like this effective for X taking finite many values.

Definition 1.5.1.

The cumulative distribution function (cdf) of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P_X(X \le x), \qquad x \in \mathscr{R}$$

Even if X may not take all real numbers, the cdf is defined for all $x \in \mathscr{R}$

Induced probability

The induced probability of X is

$$P_X(B) = P(X \in B) = P(\{\omega \in S : X(\omega) \in B\})$$

The probability P_X is called the distribution of X.

Example 1.4.3 (continued)

If the coin is fair, then

x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

A table like this effective for X taking finite many values.

Definition 1.5.1.

The cumulative distribution function (cdf) of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P_X(X \leq x), \qquad x \in \mathscr{R}$$

Even if X may not take all real numbers, the cdf is defined for all $x \in \mathscr{R}$

Induced probability

The induced probability of X is

$$P_X(B) = P(X \in B) = P(\{\omega \in S : X(\omega) \in B\})$$

The probability P_X is called the distribution of X.

Example 1.4.3 (continued)

If the coin is fair, then

x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

A table like this effective for X taking finite many values.

Definition 1.5.1.

The cumulative distribution function (cdf) of a random variable X, denoted by $F_X(x)$, is defined by

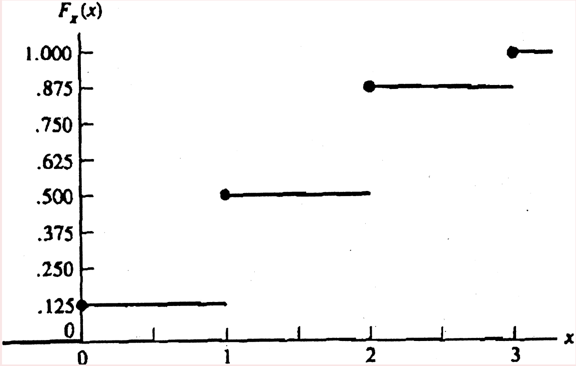
$$F_X(x) = P_X(X \leq x), \qquad x \in \mathscr{R}$$

Even if X may not take all real numbers, the cdf is defined for all $x \in \mathscr{R}$

Example 1.4.3 (continued)

The cdf of X is a step function

$$F_X(x) = \begin{cases} 0 & -\infty < x < 0 \\ 0.125 & 0 \le x < 1 \\ 0.5 & 1 \le x < 2 \\ 0.875 & 2 \le x < 3 \\ 1 & 3 \le x < \infty \end{cases}$$



Theorem 1.5.3.

The function $F(x)$ is a cdf iff

a. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$;

b. $F(x)$ is nondecreasing in x;

c. $F(x)$ is right-continuous: $\lim_{\varepsilon\to 0, \varepsilon>0} F(x+\varepsilon) = F(x)$ for any $x \in \mathscr{R}$.

This theorem says that

- if F is the cdf of a random variable X, then F satisfies a-c (this is easy to prove);
- if F satisfies a-c, then there exists a random variable X such that the cdf of X is F (this is not easy to prove).

Definition 1.5.7.

A random variable X is continuous if $F_X(x)$ is continuous in x.
A random variable X is discrete if $F_X(x)$ is a step function of x.

There are random variables that are neither discrete nor continuous; e.g., a mixture of the two types.

Theorem 1.5.3.

The function $F(x)$ is a cdf iff
a. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$;
b. $F(x)$ is nondecreasing in x;
c. $F(x)$ is right-continuous: $\lim_{\varepsilon \to 0, \varepsilon > 0} F(x+\varepsilon) = F(x)$ for any $x \in \mathscr{R}$.

This theorem says that

- if F is the cdf of a random variable X, then F satisfies a-c (this is easy to prove);
- if F satisfies a-c, then there exists a random variable X such that the cdf of X is F (this is not easy to prove).

Definition 1.5.7.

A random variable X is continuous if $F_X(x)$ is continuous in x.
A random variable X is discrete if $F_X(x)$ is a step function of x.

There are random variables that are neither discrete nor continuous; e.g., a mixture of the two types.

Theorem 1.5.3.

The function $F(x)$ is a cdf iff
a. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$;
b. $F(x)$ is nondecreasing in x;
c. $F(x)$ is right-continuous: $\lim_{\varepsilon\to 0,\varepsilon>0} F(x+\varepsilon) = F(x)$ for any $x \in \mathscr{R}$.

This theorem says that

- if F is the cdf of a random variable X, then F satisfies a-c (this is easy to prove);
- if F satisfies a-c, then there exists a random variable X such that the cdf of X is F (this is not easy to prove).

Definition 1.5.7.

A random variable X is continuous if $F_X(x)$ is continuous in x.
A random variable X is discrete if $F_X(x)$ is a step function of x.

There are random variables that are neither discrete nor continuous; e.g., a mixture of the two types.

Example 1.5.5 (logistic distribution)

An example of a continuous cdf is

$$F(x) = \frac{1}{1+e^{-x}}, \qquad x \in \mathscr{R}$$

It satisfies a-c in Theorem 1.5.3, because

$$\lim_{x\to-\infty} \frac{1}{1+e^{-x}} = 0, \qquad \lim_{x\to\infty} \frac{1}{1+e^{-x}} = 1$$

F is continuous everywhere; in fact, it is differentiable everywhere with

$$\frac{d}{dx}F(x) = F'(x) = \frac{e^{-x}}{(1+e^{-x})^2} > 0 \qquad x \in \mathscr{R}$$

Hence, F is strictly increasing.

Example

Let c be a constant and

$$G_c(x) = \begin{cases} 0 & x < 0 \\ 1 - ce^{-3x} & x \geq 0 \end{cases}$$

For what values of c, G_c is a cdf?

Example 1.5.5 (logistic distribution)

An example of a continuous cdf is

$$F(x) = \frac{1}{1+e^{-x}}, \qquad x \in \mathscr{R}$$

It satisfies a-c in Theorem 1.5.3, because

$$\lim_{x\to-\infty} \frac{1}{1+e^{-x}} = 0, \qquad \lim_{x\to\infty} \frac{1}{1+e^{-x}} = 1$$

F is continuous everywhere; in fact, it is differentiable everywhere with

$$\frac{d}{dx}F(x) = F'(x) = \frac{e^{-x}}{(1+e^{-x})^2} > 0 \qquad x \in \mathscr{R}$$

Hence, F is strictly increasing.

Example

Let c be a constant and

$$G_c(x) = \begin{cases} 0 & x < 0 \\ 1 - ce^{-3x} & x \geq 0 \end{cases}$$

For what values of c, G_c is a cdf?

Clearly,

$$\lim_{x\to-\infty} G_c(x) = 0, \qquad \lim_{x\to\infty} G_c(x) = 1$$

$$G_c'(x) = \begin{cases} 0 & x < 0 \\ 3ce^{-3x} & x > 0 \end{cases}$$

- If $c < 0$, G_c is decreasing.
- If $c > 1$, $G_c(0) = 1 - c < 0$.
- If $0 \le c \le 1$, then G_c is nondecreasing.
- G_c is always right-continuous.

Hence, G_c is a cdf iff $0 \le c \le 1$.
If $c = 1$, then G_c is continuous.
If $c = 0$, then G_c is a special discrete cdf.
If $0 < c < 1$, then G_c is neither continuous nor discrete.

Definition 1.5.8/Theorem 1.5.10.

Two random variables X and Y with cdf's F_X and F_Y respectively are identically distributed iff $F_X(x) = F_Y(x)$ for any $x \in \mathscr{R}$, which is equivalent to $P(X \in B) = P(Y \in B)$ for any Borel set B.

Clearly,

$$\lim_{x\to-\infty} G_c(x) = 0, \qquad \lim_{x\to\infty} G_c(x) = 1$$

$$G_c'(x) = \begin{cases} 0 & x < 0 \\ 3ce^{-3x} & x > 0 \end{cases}$$

- If $c < 0$, G_c is decreasing.
- If $c > 1$, $G_c(0) = 1 - c < 0$.
- If $0 \le c \le 1$, then G_c is nondecreasing.
- G_c is always right-continuous.

Hence, G_c is a cdf iff $0 \le c \le 1$.
If $c = 1$, then G_c is continuous.
If $c = 0$, then G_c is a special discrete cdf.
If $0 < c < 1$, then G_c is neither continuous nor discrete.

Definition 1.5.8/Theorem 1.5.10.

Two random variables X and Y with cdf's F_X and F_Y respectively are identically distributed iff $F_X(x) = F_Y(x)$ for any $x \in \mathscr{R}$, which is equivalent to $P(X \in B) = P(Y \in B)$ for any Borel set B.

Note that two identically distributed random variables are not necessarily equal (see Example 1.5.9).

The cdf's can be used to calculate various probabilities related to random variables but sometimes it is more convenient to use another function to calculate probabilities.

Definition 1.6.1.

The probability mass function (pmf) of a discrete random variable X is

$$f_X(x) = P(X = x) \qquad x \in \mathscr{R}$$

- For discrete X, $f_X(x) > 0$ for only countably many x's.
- $f_X(x)$ is the size of the jump in the cdf F_X at x.
- For any Borel set A,

$$P(X \in A) = \sum_{k \in A} f_X(k) \quad \text{and} \quad F_X(x) = P(X \le x) = \sum_{k \le b} f_X(k)$$

- How to obtain the pmf for a discrete random variable X?
  - List all values X can take (with positive probabilities), $x_1, x_2, \ldots$
  - Find $P(X = x_i)$, $i = 1, 2, \ldots$

Note that two identically distributed random variables are not necessarily equal (see Example 1.5.9).

The cdf's can be used to calculate various probabilities related to random variables but sometimes it is more convenient to use another function to calculate probabilities.

Definition 1.6.1.

The probability mass function (pmf) of a discrete random variable X is

$$f_X(x) = P(X = x) \qquad x \in \mathscr{R}$$

- For discrete X, $f_X(x) > 0$ for only countably many x's.
- $f_X(x)$ is the size of the jump in the cdf F_X at x.
- For any Borel set A,

$$P(X \in A) = \sum_{k \in A} f_X(k) \quad \text{and} \quad F_X(x) = P(X \le x) = \sum_{k \le b} f_X(k)$$

- How to obtain the pmf for a discrete random variable X?
  - List all values X can take (with positive probabilities), $x_1, x_2, \ldots$
  - Find $P(X = x_i)$, $i = 1, 2, \ldots$

Note that two identically distributed random variables are not necessarily equal (see Example 1.5.9).

The cdf's can be used to calculate various probabilities related to random variables but sometimes it is more convenient to use another function to calculate probabilities.

Definition 1.6.1.

The probability mass function (pmf) of a discrete random variable X is

$$f_X(x) = P(X = x) \qquad x \in \mathscr{R}$$

- For discrete X, $f_X(x) > 0$ for only countably many x's.
- $f_X(x)$ is the size of the jump in the cdf F_X at x.
- For any Borel set A,

$$P(X \in A) = \sum_{k \in A} f_X(k) \quad \text{and} \quad F_X(x) = P(X \leq x) = \sum_{k \leq b} f_X(k)$$

- How to obtain the pmf for a discrete random variable X?
  - List all values X can take (with positive probabilities), $x_1, x_2, \ldots$
  - Find $P(X = x_i)$, $i = 1, 2, \ldots$

Characteristics of a pmf

- For any pmf f, $f(x) \geq 0$, $x \in \mathscr{R}$ and there is $\{x_1, x_2, ...\}$ such that $f(x_1) + f(x_2) + \cdots = 1$.
- If $p_i > 0$, $i = 1, 2, ...$, satisfy $p_1 + p_2 + \cdots = 1$, then $f(x_i) = p_i$, $i = 1, 2, ...$, is a pmf, where $x_1, x_2, ...$ can be arbitrary.

How to define something similar to a pmf for a continuous X?

Defining $f_X(x) = P(X = x)$ does not work for a continuous X.
For any $\varepsilon > 0$,

$$P(X = x) \leq P(x - \varepsilon < X \leq x) = F_X(x) - F_X(x - \varepsilon)$$

Since F_X is continuous at x, we obtain

$$P(X = x) \leq \lim_{\varepsilon \to 0} [F_X(x) - F_X(x - \varepsilon)] = 0$$

Hence, $P(X = x) = 0$ for any $x \in \mathscr{R}$ (compare this to a discrete X).
For a discrete X, we have

$$F_X(x) = \sum_{k \leq x} f_X(k)$$

An analog for the continuous case is to replace the sum by an integral.

Characteristics of a pmf

- For any pmf f, $f(x) \geq 0$, $x \in \mathscr{R}$ and there is $\{x_1, x_2, ...\}$ such that $f(x_1) + f(x_2) + \cdots = 1$.
- If $p_i > 0$, $i = 1, 2, ...$, satisfy $p_1 + p_2 + \cdots = 1$, then $f(x_i) = p_i$, $i = 1, 2, ...$, is a pmf, where $x_1, x_2, ...$ can be arbitrary.

How to define something similar to a pmf for a continuous X?

Defining $f_X(x) = P(X = x)$ does not work for a continuous X.
For any $\varepsilon > 0$,

$$P(X = x) \leq P(x - \varepsilon < X \leq x) = F_X(x) - F_X(x - \varepsilon)$$

Since F_X is continuous at x, we obtain

$$P(X = x) \leq \lim_{\varepsilon \to 0} [F_X(x) - F_X(x - \varepsilon)] = 0$$

Hence, $P(X = x) = 0$ for any $x \in \mathscr{R}$ (compare this to a discrete X).
For a discrete X, we have

$$F_X(x) = \sum_{k \leq x} f_X(k)$$

An analog for the continuous case is to replace the sum by an integral.

Definition 1.6.3.

The probability density function (pdf) of a continuous random variable X (if it exists) is the function $f_X(x)$ satisfying

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \qquad x \in \mathscr{R}$$

- Unlike the discrete case, a pdf of a continuous X may not exist.
- Unlike the cdf, f_X is not unique.
- For a continuous X, it is convenient to use the pdf to calculate probabilities
- If F_X is differentiable, then $f_X(x) = F'_X(x) = \frac{d}{dx}F_X(x)$.
- A continuous random variable has a pdf iff its cdf is absolutely continuous.
- If f is a pdf, the set $\{x : f(x) > 0\}$ is called its support.

Theorem 1.6.5.

A function $f(x)$ is a pdf iff

a. $f(x) \geq 0$ for all x;

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

Definition 1.6.3.

The probability density function (pdf) of a continuous random variable X (if it exists) is the function $f_X(x)$ satisfying

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \qquad x \in \mathscr{R}$$

- Unlike the discrete case, a pdf of a continuous X may not exist.
- Unlike the cdf, f_X is not unique.
- For a continuous X, it is convenient to use the pdf to calculate probabilities
- If F_X is differentiable, then $f_X(x) = F'_X(x) = \frac{d}{dx}F_X(x)$.
- A continuous random variable has a pdf iff its cdf is absolutely continuous.
- If f is a pdf, the set $\{x : f(x) > 0\}$ is called its support.

Theorem 1.6.5.

A function $f(x)$ is a pdf iff

a. $f(x) \geq 0$ for all x;

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

Definition 1.6.3.

The probability density function (pdf) of a continuous random variable X (if it exists) is the function $f_X(x)$ satisfying

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \qquad x \in \mathscr{R}$$

- Unlike the discrete case, a pdf of a continuous X may not exist.
- Unlike the cdf, f_X is not unique.
- For a continuous X, it is convenient to use the pdf to calculate probabilities
- If F_X is differentiable, then $f_X(x) = F'_X(x) = \frac{d}{dx}F_X(x)$.
- A continuous random variable has a pdf iff its cdf is absolutely continuous.
- If f is a pdf, the set $\{x : f(x) > 0\}$ is called its support.

Theorem 1.6.5.

A function $f(x)$ is a pdf iff

a. $f(x) \geq 0$ for all x;

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

Calculation of probabilities

For a continuous X with pdf f_X,

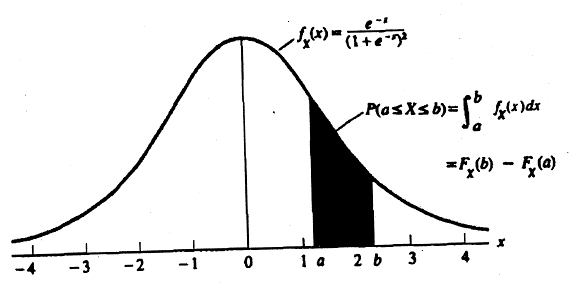
$$P(X \in A) = \int_A f_X(x)dx$$

For example,

$$P(a < X < b) = P(a < X \le b) = P(a \le X < b)$$

$$= P(a \le X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(t)dt$$

which is the area under the curve $f_X(t)$.



Example (finding the cdf, given a pdf)

Suppose that

$$f_X(x) = \begin{cases} 1+x & -1 \le x < 0 \\ 1-x & 0 \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$F_X(x) = ?$

Obviously, $F_X(x) = 0$ when $x < -1$.

For $-1 \le x < 0$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-1}^{x} (1+t)dt = \frac{t^2}{2} + t \Big|_{t=-1}^{x} = \frac{x^2}{2} + x + \frac{1}{2}$$

For $0 \le x < 1$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-1}^{0} (1+t)dt + \int_{0}^{x} (1-t)dt = \frac{1}{2} - \frac{x^2}{2} + x$$

For $x \ge 1$,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-1}^{0} (1+t)dt + \int_{0}^{1} (1-t)dt = \frac{1}{2} + \frac{1}{2} = 1$$

Finding the pdf, given a cdf

- If F_X is differentiable everywhere, then $f_X = F_X'$.
- What if F_X is differentiable except for some (at most countably many) points?
  $f_X(x) = F_X'(x)$ for x at which F_X is differentiable;
  $f_X(x)$ can be any $c \geq 0$ for x at which F_X is not differentiable.

Example

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \leq x < 1/2 \\ x^2 & 1/2 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

$f_X(x) = ?$
F_X is differentiable except at 0, 1/2 and 1.

$$f_X(x) = \begin{cases} 0 & x < 0 \\ 1/2 & 0 \leq x < 1/2 \\ 2x & 1/2 \leq x < 1 \\ 0 & x \geq 1 \end{cases}$$

Finding the pdf, given a cdf

- If F_X is differentiable everywhere, then $f_X = F'_X$.
- What if F_X is differentiable except for some (at most countably many) points?
  $f_X(x) = F'_X(x)$ for x at which F_X is differentiable;
  $f_X(x)$ can be any $c \geq 0$ for x at which F_X is not differentiable.

Example

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \leq x < 1/2 \\ x^2 & 1/2 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

$f_X(x) = ?$
F_X is differentiable except at 0, 1/2 and 1.

$$f_X(x) = \begin{cases} 0 & x < 0 \\ 1/2 & 0 \leq x < 1/2 \\ 2x & 1/2 \leq x < 1 \\ 0 & x \geq 1 \end{cases}$$

Example.

What values of θ and β will make $f(x)$ a pdf?

$$f(x) = \beta e^{\theta x - |x|} \qquad x \in \mathscr{R}$$

If $\theta \geq 1$, then $\lim_{x\to\infty} e^{\theta x - |x|} = \infty$; if $\theta \leq -1$, then $\lim_{x\to-\infty} e^{\theta x - |x|} = \infty$; hence, θ has to be in $(-1,1)$.

For $\theta \in (-1,1)$,

$$\int_{-\infty}^{\infty} e^{\theta x - |x|} dx = \int_{-\infty}^{0} e^{\theta x + x} dx + \int_{0}^{\infty} e^{\theta x - x} dx = \frac{1}{1+\theta} + \frac{1}{1-\theta} = \frac{2}{1-\theta^2}$$

Hence, in orde to have a pdf, we must have $\beta = (1-\theta^2)/2$.

What is the cdf of f?

For $x \leq 0$,

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1-\theta^2}{2} \int_{-\infty}^{x} e^{\theta t + t} dt = \frac{1-\theta}{2} e^{(1+\theta)x}$$

For $x > 0$,

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1-\theta^2}{2} \left[\int_{-\infty}^{0} e^{\theta t + t} dt + \int_{0}^{x} e^{\theta t - t} dt \right] = 1 - \frac{1+\theta}{2} e^{(1-\theta)x}$$