Continuous Probability

1. Examples
2. Events
3. Continuous Random Variables
4. Expectation
5. Bayes’ Rule
6. Multiple Random Variables
Continuous Probability - Pick a real number.

Choose a real number $X$, uniformly at random in $[0, 1000]$. What is the probability that $X$ is exactly equal to $100\pi = 314.1592625 \ldots$? Well, ..., 0.

Let $[a, b]$ denote the event that the point $X$ is in the interval $[a, b]$.

$$Pr([a, b]) = \frac{\text{length of } [a, b]}{\text{length of } [0, L]} = \frac{b-a}{L} = \frac{b-a}{1000}.$$ 

Intervals like $[a, b] \subseteq \Omega = [0, L]$ are events. More generally, events in this space are unions of intervals. Example: the event $A$ - “within 50 of 0 or 1000” is $A = [0, 50] \cup [950, 1000]$. Thus,

$$Pr[A] = Pr[[0, 50]] + Pr[[950, 10000]] = \frac{1}{10}.$$
Continuous Probability - Pick a random real number.

Note: A **radical** change in approach. For a finite probability space, \( \Omega = \{1, 2, \ldots, N\} \), we started with \( Pr[\omega] = p_\omega \). We then defined \( Pr[A] = \sum_{\omega \in A} p_\omega \) for \( A \subset \Omega \). We used the same approach for countable \( \Omega \).

For a continuous space, e.g., \( \Omega = [0, L] \), we cannot start with \( Pr[\omega] \), because this will typically be 0. Instead, we start with \( Pr[A] \) for some events \( A \). Here, we started with \( A = \text{interval, or union of intervals} \).

Thus, the probability is a function from events to \([0, 1]\). Can any function make sense? No! At least, it should be additive!. In our example, \( Pr[[0, 50] \cup [950, 1000]] = Pr[[0, 50]] + Pr[[950, 1000]] \).
A James Bond example. In Spectre, Mr. Hinx is chasing Bond who is in a Aston Martin DB10. Hinx shoots at the DB10 and hits it at a random spot. What is the chance Hinx hits the gas tank? Assume the gas tank is a one foot circle and the DB10 is an expensive $4 \times 5$ rectangle.

\[
\Omega = \{(x, y) : x \in [0, 4], y \in [0, 5]\}.
\]

The size of the event is $\pi(1)^2 = \pi$.

The “size” of the sample space which is $4 \times 5$.

Since uniform, probability of event is $\frac{\pi}{20}$.
Continuous Random Variables: CDF

\[ Pr[a < X \leq b] \] instead of \[ Pr[X = a] \].
For all \(a\) and \(b\): specifies the behavior!
Simpler: \( P[X \leq x] \) for all \(x\).

**Cumulative probability Distribution Function** of \(X\) is

\[
F_X(x) = Pr[X \leq x] \tag{1} \]

\[ Pr[a < X \leq b] = Pr[X \leq b] - Pr[X \leq a] = F_X(b) - F_X(a). \]

Idea: two events \(X \leq b\) and \(X \leq a\).
Difference is the event \(a < X \leq b\).
Indeed: \( \{X \leq b\} \setminus \{X \leq a\} = \{X \leq b\} \cap \{X > a\} = \{a < X \leq b\} \).
Example: CDF

Example: Value of $X$ in $[0, L]$ with $L = 1000$.

$$F_X(x) = Pr[X \leq x] = \begin{cases} 
0 & \text{for } x < 0 \\
\frac{x}{1000} & \text{for } 0 \leq x \leq 1000 \\
1 & \text{for } x > 1000 
\end{cases}$$

Probability that $X$ is within 50 of center:

$$Pr[450 < X \leq 550] = Pr[X \leq 550] - Pr[X \leq 450]$$

$$= \frac{550}{1000} - \frac{450}{1000}$$

$$= \frac{100}{1000} = \frac{1}{10}$$
Example: CDF

Example: hitting random location on gas tank. Random location on circle.

Random Variable: \( Y \) distance from center. Probability within \( y \) of center:

\[
P_r[Y \leq y] = \frac{\text{area of small circle}}{\text{area of dartboard}} = \frac{\pi y^2}{\pi} = y^2.
\]

Hence,

\[
F_Y(y) = Pr[Y \leq y] = \begin{cases} 
0 & \text{for } y < 0 \\
\frac{y^2}{\pi} & \text{for } 0 \leq y \leq 1 \\
1 & \text{for } y > 1
\end{cases}
\]
Calculation of event with dartboard.

Probability between .5 and .6 of center?
Recall CDF.

\[ F_Y(y) = Pr[Y \leq y] = \begin{cases} 
0 & \text{for } y < 0 \\
y^2 & \text{for } 0 \leq y \leq 1 \\
1 & \text{for } y > 1 
\end{cases} \]

\[ Pr[0.5 < Y \leq 0.6] = Pr[Y \leq 0.6] - Pr[Y \leq 0.5] \]
\[ = F_Y(0.6) - F_Y(0.5) \]
\[ = 0.36 - 0.25 \]
\[ = 0.11 \]
Density function.

Is the dart more like to be (near) .5 or .1? Probability of “Near x” is \( Pr[x < X \leq x + \delta] \).
Goes to 0 as \( \delta \) goes to zero.
Try

\[
Pr[x < X \leq x + \delta] \quad \frac{\delta}{\delta}
\]

The limit as \( \delta \) goes to zero.

\[
\lim_{\delta \to 0} \frac{Pr[x < X \leq x + \delta]}{\delta} = \lim_{\delta \to 0} \frac{Pr[X \leq x + \delta] - Pr[X \leq x]}{\delta}
\]

\[
= \lim_{\delta \to 0} \frac{F_X(x + \delta) - F_X(x)}{\delta}
\]

\[
= \frac{d(F_X(x))}{dx}.
\]
Definition: (Density) A probability density function for a random variable \( X \) with cdf \( F_X(x) = Pr[X \leq x] \) is the function \( f_X(x) \) where

\[
F_X(x) = \int_{-\infty}^{x} f_X(u) du.
\]

Thus,

\[
Pr[X \in (x, x + \delta)] = F_X(x + \delta) - F_X(x) \approx f_X(x) \delta.
\]
Examples: Density.

Example: uniform over interval \([0, 1000]\)

\[
f_X(x) = F'_X(x) = \begin{cases} 
  0 & \text{for } x < 0 \\
  \frac{1}{1000} & \text{for } 0 \leq x \leq 1000 \\
  0 & \text{for } x > 1000 
\end{cases}
\]

Example: uniform over interval \([0, L]\)

\[
f_X(x) = F'_X(x) = \begin{cases} 
  0 & \text{for } x < 0 \\
  \frac{1}{L} & \text{for } 0 \leq x \leq L \\
  0 & \text{for } x > L 
\end{cases}
\]
Examples: Density.

Example: “Dart” board.
Recall that

\[ F_Y(y) = Pr[Y \leq y] = \begin{cases} 
0 & \text{for } y < 0 \\
y^2 & \text{for } 0 \leq y \leq 1 \\
1 & \text{for } y > 1
\end{cases} \]

\[ f_Y(y) = F'_Y(y) = \begin{cases} 
0 & \text{for } y < 0 \\
2y & \text{for } 0 \leq y \leq 1 \\
0 & \text{for } y > 1
\end{cases} \]

The cumulative distribution function (cdf) and probability distribution function (pdf) give full information. Use whichever is convenient.
Target

Random Variable

Event \{Y \leq y\}

Outcome

$F_Y(y)$

$y^2$

$1$ $1$

$2y$

$2$ $1$
Uniform in $[a, b]$
The exponential distribution with parameter $\lambda > 0$ is defined by

$$f_X(x) = \lambda e^{-\lambda x} 1\{x \geq 0\}$$

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0. \end{cases}$$

Note that $Pr[X > t] = e^{-\lambda t}$ for $t > 0$. 
Random Variables

Continuous random variable $X$, specified by

1. $F_X(x) = Pr[X \leq x]$ for all $x$.
   
   **Cumulative Distribution Function (cdf).**
   
   $Pr[a < X \leq b] = F_X(b) - F_X(a)$
   
   1.1 $0 \leq F_X(x) \leq 1$ for all $x \in \mathbb{R}$.
   1.2 $F_X(x) \leq F_X(y)$ if $x \leq y$.

2. Or $f_X(x)$, where $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ or $f_X(x) = \frac{d(F_X(x))}{dx}$.
   
   **Probability Density Function (pdf).**
   
   $Pr[a < X \leq b] = \int_{a}^{b} f_X(x)dx = F_X(b) - F_X(a)$
   
   2.1 $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.
   2.2 $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Recall that $Pr[X \in (x, x + \delta)] \approx f_X(x)\delta$. Think of $X$ taking discrete values $n\delta$ for $n = \ldots, -2, -1, 0, 1, 2, \ldots$ with $Pr[X = n\delta] = f_X(n\delta)\delta$. 
The pdf $f_X(x)$ is a nonnegative function that integrates to 1.

The cdf $F_X(x)$ is the integral of $f_X$.

$$Pr[x < X < x + \delta] \approx f_X(x)\delta$$

$$Pr[X \leq x] = F_X(x) = \int_{-\infty}^{x} f_X(u)du$$
Some Examples

1. **Expo is memoryless.** Let \( X = Expo(\lambda) \). Then, for \( s, t > 0 \),

\[
Pr[X > t + s \mid X > s] = \frac{Pr[X > t + s]}{Pr[X > s]}
= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t}
= Pr[X > t].
\]

‘Used is a good as new.’

2. **Scaling Expo.** Let \( X = Expo(\lambda) \) and \( Y = aX \) for some \( a > 0 \). Then

\[
Pr[Y > t] = Pr[aX > t] = Pr[X > t/a]
= e^{-\lambda(t/a)} = e^{-(\lambda/a)t} = Pr[Z > t] \text{ for } Z = Expo(\lambda/a).
\]

Thus, \( a \times Expo(\lambda) = Expo(\lambda/a) \).
Some More Examples

3. Scaling Uniform. Let $X = U[0, 1]$ and $Y = a + bX$ where $b > 0$. Then,

$$Pr[Y \in (y, y + \delta)] = Pr[a + bX \in (y, y + \delta)] = Pr[X \in (\frac{y-a}{b}, \frac{y+\delta-a}{b})]$$

$$= Pr[X \in (\frac{y-a}{b}, \frac{y-a}{b} + \frac{\delta}{b})] = \frac{1}{b} \delta, \text{ for } 0 < \frac{y-a}{b} < 1$$

$$= \frac{1}{b} \delta, \text{ for } a < y < a+b.$$  

Thus, $f_Y(y) = \frac{1}{b}$ for $a < y < a+b$. Hence, $Y = U[a, a+b]$.

4. Scaling pdf. Let $f_X(x)$ be the pdf of $X$ and $Y = a + bX$ where $b > 0$. Then

$$Pr[Y \in (y, y + \delta)] = Pr[a + bX \in (y, y + \delta)] = Pr[X \in (\frac{y-a}{b}, \frac{y+\delta-a}{b})]$$

$$= Pr[Pr[X \in (\frac{y-a}{b}, \frac{y-a}{b} + \frac{\delta}{b})] = f_X(\frac{y-a}{b}) \frac{\delta}{b}.$$  

Now, the left-hand side is $f_Y(y)\delta$. Hence,

$$f_Y(y) = \frac{1}{b} f_X(\frac{y-a}{b}).$$
**Expectation**

**Definition** The expectation of a random variable $X$ with pdf $f(x)$ is defined as

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

Justification: Say $X = n\delta$ w.p. $f_X(n\delta)\delta$. Then,

$$E[X] = \sum_n (n\delta) Pr[X = n\delta] = \sum_n (n\delta)f_X(n\delta)\delta = \int_{-\infty}^{\infty} xf_X(x)dx.$$  

Indeed, for any $g$, one has $\int g(x)dx \approx \sum_n g(n\delta)\delta$. Choose $g(x) = xf_X(x)$. 

![Diagram showing expectation and pdf](image)
**Definition** The expectation of a function of a random variable is defined as

\[ E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) \, dx. \]

Justification: Say \( X = n\delta \) w.p. \( f_X(n\delta)\delta \). Then,

\[ E[h(X)] = \sum_n h(n\delta) \Pr[X = n\delta] = \sum_n h(n\delta) f_X(n\delta)\delta = \int_{-\infty}^{\infty} h(x) f_X(x) \, dx. \]

Indeed, for any \( g \), one has \( \int g(x) \, dx \approx \sum_n g(n\delta)\delta \). Choose \( g(x) = h(x) f_X(x) \).

**Fact** Expectation is linear. **Proof**: As in the discrete case.
**Definition:** The **variance** of a continuous random variable $X$ is defined as

$$\text{var}[X] = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

$$= \int_{-\infty}^{\infty} x^2 f(x) \, dx - \left( \int_{-\infty}^{\infty} x f(x) \, dx \right)^2.$$
Motivation for Gaussian Distribution

Key fact: The sum of many small independent RVs has a Gaussian distribution.

This is the Central Limit Theorem. (See later.)

Examples: Binomial and Poisson suitably scaled.

This explains why the Gaussian distribution (the bell curve) shows up everywhere.
1. pdf: \( Pr[X \in (x, x + \delta)] = f_X(x)\delta. \)
2. CDF: \( Pr[X \leq x] = F_X(x) = \int_x^{-\infty} f_X(y)dy. \)
3. \( U[a, b], \text{Expo}(\lambda), \text{target}. \)
4. Expectation: \( E[X] = \int_{-\infty}^{\infty} xf_X(x)dx. \)
5. Expectation of function: \( E[h(X)] = \int_{-\infty}^{\infty} h(x)f_X(x)dx. \)
7. Gaussian: \( \mathcal{N}(\mu, \sigma^2) : f_X(x) = ... \) “bell curve”