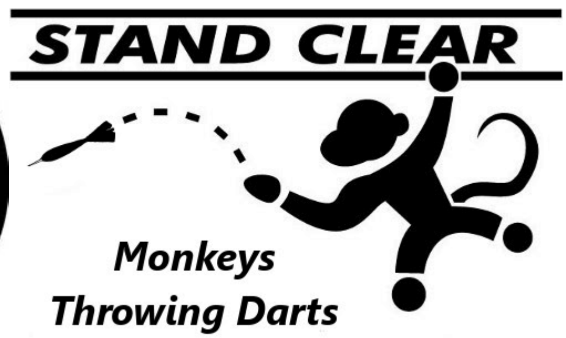
Section 4.1, 4.2: Probability Distributions for a Continuous Random Variable

A continuous random variable contains an uncountably infinite number of different values making it impossible to list each one. A different approach for a probability model is needed. Consider an unskilled player throwing darts at a board. Different regions of different sizes exist on the board. The likelihood of an unskilled player hitting the center white ring is not the same as hitting anywhere in the 20 region. How might we approach the assignment of probabilities for the dartboard?



Let's assume a dart always hits in the scoring area (which is a big assumption) and all points of contact on the board equally likely. We now let the relative area of a region on the board serve as the probability that a dart lands in said region.

Definition 1 *A probability distribution for a continuous random variable X is given by a* ***probability density function (pdf)*** *$f(x)$. The probability that X takes a value in the interval $[a,b]$ is the area under the curve $f(x)$ from a to b. Every pdf satisfies two properties.*

1. *The total area under the curve defined by $f(x)$ is 1;*

2. *and $f(x) \geq 0$ for all x.*

Example 2 *Is $f(x) = x$ on the domain $[0,1]$ a pdf? Note that we can easily compute area under this curve geometric formulae or integration techniques. No, because the area in the triangle generated by the curve is $\frac{1*1}{2} = \frac{1}{2} < 1$. Alternatively, no since $\int_0^1 x dx = \frac{x^2}{2}|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$.*

Example 3 *Is $f(x) = x$ on the domain $[0,3]$ a pdf? No, because the area in the triangle generated by the curve is $\frac{3*3}{2} = \frac{9}{2} > 1$*

Example 4 *Is $f(x) = \frac{x}{18}$ on the domain $[0,6]$ a pdf? Yes, since the area in the triangle generated by the curve is $\frac{6*\frac{6}{18}}{2} = 1$ and $f(x) \geq 0$ for all $x \in [0,6]$. .*

Problem 5 *Is* $f(x) = x$ *on the domain* $[-1, 2]$ *a pdf?*

Problem 6 *Is* $f(x) = x$ *on the domain* $[2, 3]$ *a pdf?*

Example 7 *For the pdf* $f(x) = \frac{x}{18}$ *with domain* $[0, 6]$ *, what is the probability that an observations falls in the interval* $[1, 3]$*? The area in* $[0, 3]$ *is* $\frac{3 * \frac{3}{18}}{2} = \frac{1}{4}$*. The area in* $[0, 1]$ *is* $\frac{1 * \frac{1}{18}}{2} = \frac{1}{36}$*. The* $P(x \in [1, 3]) = \frac{1}{4} - \frac{1}{36} = \frac{2}{9}$*.*

Example 8 *For the pdf* $f(x) = \frac{x}{18}$ *with domain* $[0, 6]$ *, what is the probability that* $x = 5$*? Note that* $P(x = 5) = 0$ *since the line at* $x = 5$ *is one dimensional and has no area.*

Problem 9 *For the pdf* $f(x) = \frac{x}{18}$ *with domain* $[0, 6]$ *, what is the probability that an observations falls in the interval* $[2, 5]$*?*

Problem 10 *For the pdf* $f(x) = \frac{x}{18}$ *with domain* $[0, 6]$ *, what is the probability that an observations falls in the interval* $[5, 6]$*?*

Example 11 *Let* $f(x) = cx$ *on the domain* $[0, 10]$*. Find the constant* c *that makes* $f(x)$ *a pdf. We know that*

$$\begin{aligned} 1 &= \int_0^{10} cx dx \\ &= c \int_0^{10} x dx \\ &= c * \left(\frac{x^2}{2} |_0^{10} \right) \\ &= c * 50. \end{aligned}$$

Thus, $c = \frac{1}{50}$*.*

Example 12 *Let $f(x) = cx^2$ on the domain $[0, 5]$. Find the constant c that makes $f(x)$ a pdf. We know that*

$$\begin{aligned} 1 &= \int_0^5 cx^2 dx \\ &= c\int_0^5 x^2 dx \\ &= c\left(\frac{x^3}{3}|_0^5\right) \\ &= c * \frac{125}{3}. \end{aligned}$$

So, $c = \frac{3}{125}$.

Definition 13 *The **uniform distribution** is distributed uniformly between two points a and b. In a uniform distribution the pdf $f(x) = \frac{1}{b-a}$ on the domain $[a, b]$.*

Note that $f(x) = \frac{1}{b-a}$ is a horizontal line. The area under $f(x) = \frac{1}{b-a}$ on any interval forms a rectangle.

Example 14 *Consider the uniform distribution the interval $[6, 10]$. Compute $P(x \in [7, 8.5])$. Note that $f(x) = \frac{1}{4}$.*
The area of the rectangle from 7 to 8.5 under the curve $f(x)$ is $1.5 * \frac{1}{4} = 0.375$. So the probability that a randomly selected observation falls into the interval $[7, 8.5]$ is 0.375. Alternatively $P(x \in [7, 8.5]) = \int_7^{8.5} \frac{dx}{4} = \frac{x}{4}|_7^{8.5} = 0.375$.

Problem 15 *What happens if instead of the closed interval $[7, 8.5]$, we ask about $P(x \in (7, 8.5))$? In other words, do endpoints matter? They certainly do in a discrete random variable. Is the same true for a continuous random variable? Using geometry, how much area does the line $x = 8.5$ occupy under the curve $f(x) = \frac{1}{4}$? Lines are two-dimensional and have no area. So, $P(x = 8.5) = 0$. Alternatively, $P(x = 8.5) = \int_{8.5}^{8.5} \frac{dx}{4} = \frac{x}{4}|_{8.5}^{8.5} = 0.0$. In a continuous random variable the probability that x equals one specific value is 0. Hence, $P(x \in (7, 8.5)) = P(x \in [7, 8.5]) = 0.375$.*

Cumulative distribution functions in a continuous random variable behave in a similar fashion to discrete random variables. $F(x) = P(x \leq X)$ Since X is uncountable this is no longer a sum but rather an integral. $F(x) = P(x \leq X) = \int_{-\infty}^{x} f(x)dx$.

Example 16 *Consider the uniform distribution the interval $[6, 10]$. Note that*

$f(x) = \frac{1}{4}$. *Compute* $F(8.5)$.

$$\begin{aligned} F(8.5) &= P(x \leq 8.5) \\ &= \int_{-\infty}^{8.5} \frac{dx}{4} \\ &= \int_{6}^{8.5} \frac{dx}{4} \\ &= \frac{x}{4}|_{6}^{8.5} \\ &= 0.625 \end{aligned}$$

We can write $F(x)$ *as a piece-wise defined function. For any* $j \in [6, 10]$,

$$\begin{aligned} F(j) &= P(x \leq j) \\ &= \int_{-\infty}^{j} \frac{dx}{4} \\ &= \int_{6}^{j} \frac{dx}{4} \\ &= \frac{x}{4}|_{6}^{j} \\ &= \frac{j-6}{4} \end{aligned}$$

and $F(x) = \left\{ \begin{array}{cc} 0 & x < 6 \\ \frac{x-6}{4} & 6 \leq x < 10 \\ 1 & x \geq 10 \end{array} \right\}$.

Remark 17 *Note that for all* a, b

1. $P(x > a) = 1 - F(a)$;
2. $P(a \leq x \leq b) = F(b) - F(a)$.

So now we can find $P(x \in [7, 8.5])$ as $F(8.5) - F(7) = \frac{8.5-6}{4} - \frac{7-6}{4} = 0.375$

We were able to easily find $F(x)$ when given $f(x)$. Can we find $f(x)$ from $F(x)$? Yes, wherever $F(x)$ is differentiable we get that $F'(x) = f(x)$. Note that the derivative of $F(x) = \left\{ \begin{array}{cc} 0 & x < 6 \\ \frac{x-6}{4} & 6 \leq x < 10 \\ 1 & x \geq 10 \end{array} \right\}$ is defined on $(6, 10)$ and $F'(x) = \left\{ \begin{array}{cc} 0 & x < 6 \\ \frac{1}{4} & 6 \leq x < 10 \\ 0 & x \geq 10 \end{array} \right\}$.

Cumulative distribution functions make it easy to compute percentile values.

Example 18 *Consider the uniform distribution the interval* $[6, 10]$*. Find* P_{40}*.* P_{40} *is the* x *value where* $F(x) = .4$*. So, we solve for* $\dfrac{x-6}{4} = .4$ *, Solution is:* 7.6.

Example 19 *Consider the uniform distribution the interval* $[6, 10]$*. Find the median. Since med* $= P_{50}$*, we replicate the previous example. We solve for* $\dfrac{x-6}{4} = .5$*, Solution is:* 8.0.