Stat 400 section 4.2 Cumulative Distribution Functions and Expected Value

notes by Tim Pilachowski

A *probability density function* or *probability distribution function* for a continuous random variable is a function $f(x)$ such that, on an interval $[a, b]$,

$$P(a \le X \le b) = \int_a^b f(x)\,dx .$$

A pdf $f(x)$ has two necessary characteristics:

1. $f(x) \ge 0$ for all values of x in the interval $[a, b]$ (since all probabilities, and therefore areas under the curve, are zero or positive)
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (since the sum of all probabilities = 1 = area under the curve over the entire domain)

When we evaluate the integral of a probability density function for a portion of the domain from $-\infty$ to an arbitrary value of x, $P(X \le x) = \int_{-\infty}^{x} f(y)\,dy = F(x)$, we get another function which is an antiderivative and is called the *cumulative distribution function*.

Formally, given a probability density function defined as $\begin{cases} f(x) & A \le x \le B \\ 0 & \text{otherwise} \end{cases}$,

the cumulative density function will have the form $P(X \le x) = \begin{cases} 0 & x < A \\ F(x) & A \le x < B \\ 1 & B \le x \end{cases}$.

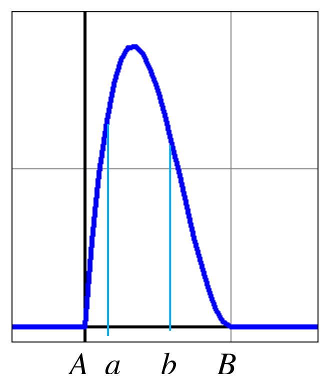
A cumulative distribution function has several useful characteristics.

Since $\int f(x)\,dx = F(x)$, we know that (where it exists) $F'(x) = f(x)$.

Since $\int_a^b f(x)\,dx = F(b) - F(a)$, we can use $F(x)$ to find probabilities directly: $P(a \le X \le b) = F(b) - F(a)$.

Since $\int_A^A f(x)\,dx = 0$, it will always be true that $F(A) = 0$.

Since $\int_A^B f(x)\,dx = F(B) - F(A) = 1 - 0$, it will always be true that $F(B) = 1$.



Example A: Given the probability density function $f(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$ find the associated cumulative density function $F(x)$ then use it to find $P\left(\frac{1}{2} \le X \le \frac{3}{4}\right)$. *answer*: $\frac{67}{256} = 0.26171875$

Example B: Given the probability density function $f(x)=\begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$ find the associated cumulative density function $F(x)$ then use it to find $P\left(X \le \frac{1}{2}\right)$ and $P\left(X \ge \frac{3}{4}\right)$. *answer*: 0.5, $\frac{5}{32} = 0.15625$

For a number $0 \le p \le 1$, the $(100p)^{\text{th}}$ *percentile* of the distribution of a continuous random variable X, denoted by $\eta(p)$ is defined by $p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)\,dy$. From a less technical standpoint, set the cumulative distribution function $F(x) = p$ and solve.

Example C: A continuous random variable X has the probability density function $f(x)=\begin{cases} 2x-2 & 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$. First determine the cumulative distribution function, then use it to find the 25^{th}, 50^{th} and 75^{th} percentiles (lower quartile, median $\tilde{\mu}$, and upper quartile). *answers*: $\frac{3}{2},\ 1+\frac{\sqrt{2}}{2},\ 1+\frac{\sqrt{3}}{2}$

Example D: A continuous random variable X has cumulative distribution function $F(x)=\begin{cases} 0 & x<0 \\ \frac{1}{27}x^3 & 0 \le x \le 3 \\ 1 & 3<x \end{cases}$.

Find its probability density function $f(x)$.

Example E: A continuous random variable X has probability density function $f(x) = \begin{cases} \frac{1}{10} & 0 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$.

Find the median. *answer*: $\tilde{\mu} = 5$

Example B revisited: Given the probability density function $f(x) = \begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$ find $\tilde{\mu}$.

answer: 0.5

Example F: Gauge 4 aluminum sheets should have a thickness of 0.2043 inches, with a tolerance of ± 0.011 inches allowed. During the rolling process that produces metal sheet stock, a certain amount of "bowing" occurs in the rollers. The quality control department of an aluminum factory believes that one of the company's rolling machines is producing sheets with thicknesses varying between 0.1931 and 0.2168 inches. If we assume thickness is a uniform random variable, what is the median thickness? *answer*: $\tilde{\mu} = 0.20495$

Example G. For a particular machine, its useful lifetime (random variable T) is modeled by $f(t) = \begin{cases} 0.1e^{-0.1t} & 0 \le x < \infty \\ 0 & \text{otherwise} \end{cases}$. Determine the median and interpret its meaning in this situation.

answer: $\tilde{\mu} = -10 \ln(0.5) = 10 \ln(2)$ years

Example H. Suppose that the number of days a patient spends in ICU can be approximated with a probability distribution function $f(t) = \begin{cases} 0.4e^{-0.4(t-1)} & t > 1 \\ 0 & \text{otherwise} \end{cases}$. Find the median and interpret.

answer: $\tilde{\mu} = -2.5 \ln(0.5) + 1 = 2.5 \ln(2) + 1$ days

Recall from section 3.3, for *discrete* random variables:

$$E(X) = \mu_X = x_1\, p(x_1) + x_2\, p(x_2) + x_3\, p(x_3) + \ldots + x_n\, p(x_n) = \sum_x x_i\, p(x_i) = \text{ sum of [value * probability]}$$

$$V(X) = \sigma_X^2 = \sigma^2 = (x_1 - \mu)^2 * p(x_1) + (x_2 - \mu)^2 * p(x_2) + \ldots + (x_n - \mu)^2 * p(x_n) = \sum_x (x_i - \mu)^2 * p(x_i)$$

and the shortcut shortcut formula for variance:

$$V(X) = \sigma_X^2 = \left[\sum_x (x_i)^2 * p(x_i)\right] - \mu^2 = E(X^2) - [E(X)]^2 .$$

If we apply the same underlying concept to continuous random variables, we get analogous integrals. Given a random variable X with a probability density function $f(x)$:

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x * f(x)\, dx \qquad\qquad V(X) = \sigma_X^2 = E(X^2) - [E(X)]^2 = \int_{-\infty}^{\infty} x^2 * f(x)\, dx - \mu^2 .$$

As before, standard deviation of $X = \sigma_X = \sqrt{V(X)}$.

Example A revisited: Given the probability density function $f(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$ find the expected value and variance. *answers*: $\frac{2}{5}$, $\frac{1}{25}$

Example B again: Given the probability density function $f(x) = \begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$ find $E(X)$ and $V(X)$.

answers: $\frac{1}{2}$, $\frac{1}{20}$

Example C revisited: Given the probability density function $f(x) = \begin{cases} 2x-2 & 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$ determine the probability that the value of X will be within one standard deviation of the mean. *answer*: ≈ 0.6285

Example C once more. Let continuous random variable $Y = 3X + 10$, where the probability density function for random variable X is $f(x) = \begin{cases} 2x-2 & 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$. Find $E(Y)$ and $V(Y)$. *answers*: 15, $\frac{1}{2}$

Example D revisited: A continuous random variable X has cumulative distribution function

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{27}x^3 & 0 \le x \le 3 \\ 1 & 3 < x \end{cases}.$$

Find $E(X)$ and $V(X)$. *answers*: $\frac{9}{4}$, $\frac{27}{80}$

Example E revisited: A continuous random variable X has probability density function $f(x) = \begin{cases} \frac{1}{10} & 0 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$.

find $E(X)$ and $V(X)$. *answers*: 5, $\frac{25}{3}$

Example F revisited: Gauge 4 aluminum sheets should have a thickness of 0.2043 inches, with a tolerance of ± 0.011 inches allowed. During the rolling process that produces metal sheet stock, a certain amount of "bowing" occurs in the rollers. The quality control department of an aluminum factory believes that one of the company's rolling machines is producing sheets with thicknesses varying between 0.1931 and 0.2168 inches. If we assume thickness is a uniform random variable, what is the expected thickness? What is the variance?
answers: 0.20495, ≈ 0.04205

Example G revisited. For a particular machine, its useful lifetime (random variable T) is modeled by $f(t)=\begin{cases} 0.1e^{-0.1t} & 0 \le x < \infty \\ 0 & \text{otherwise} \end{cases}$. Find the expected number of years that the machine will last, along with the standard deviation. *answers*: 10 years, 10

Example H revisited. Suppose that the number of days a patient spends in ICU can be approximated with a probability distribution function $f(t)=\begin{cases} 0.4e^{-0.4(t-1)} & t > 1 \\ 0 & \text{otherwise} \end{cases}$. Find the mean and interpret its meaning in this situation. What is the standard deviation? Hint: Before you start evaluating your integrals, do a substitution: $w = t - 1$. It makes the integration by parts a good bit easier. *answers*: 3.5 days, 2.5