Continuous Random Variables

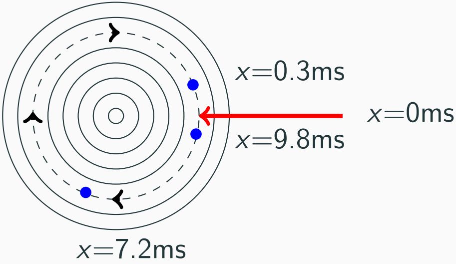
David Gerard
Most slides borrowed from Linda Collins

Learning Objectives

- Continuous random variables.
- Uniform distribution.
- Mean of continuous random variables.
- Variance of continuous random variables.
- Sections 2.5

Continuous Random Variables i

Let X = the time it takes for a read/write head to locate a specific file header on a computer hard disk drive (once the head has been positioned over the correct track).



The disk makes one complete revolution every 10 milliseconds.

Continuous Random Variables ii

We want to answer questions like:

1. What is the chance of locating the file... in less than 2ms?
2. What is the chance of locating the file... in somewhere between 5ms and 7ms?
3. What is the average wait time?
   (That is, the average time to locate the file.)

Continuous Random Variables iii

X is a random variable.

- X assigns a number to each outcome of a random experiment
- The experiment results in a random outcome because the disk is spinning. Once the read/write head reaches the correct track, the wait time until the file header comes around to meet the read/write head is random.
- Since the disk spins at a constant rate of 10ms per revolution, we should be able to determine probabilities for wait times.

What are the possible values for X?

Sample Space $= S_X = [0, 10)$, which is an interval.

That is, X can take on any real number: $0 \leq x \leq 10$.

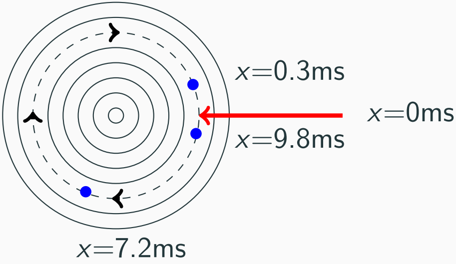
Continuous Random Variables iv

Consider the following probabilities:

- $P(2 \leq X \leq 4)$
- $P(1 \leq X \leq 3)$
- $P(5 \leq X \leq 7)$

All are intervals of length 2ms.

Is wait time X more likely to be in one 2ms interval than another?



Continuous Random Variables v

Since all 2ms intervals within [0,10) are equally likely, then

(A) $P(2 \leq X \leq 4) = P(1 \leq X \leq 3) = P(5 \leq X \leq 7)$ and we say that X has a Uniform distribution

(B) ...and for each x in S_X, $P(X = x) = c > 0$.

Wait! (B) cannot be true! ...because then $P(S_X) = \infty \neq 1$.

The sample space $S_X = [0, 10)$ is an interval, *not* a discrete set.

S_X is not countable, listable.

We cannot sum $P(X = x) = c$ for all x in [0,10).

Continuous Random Variables iv

So, X is not a discrete random variable.

We say that X is a continuous Uniform random variable, since the sample space is an interval.

- Still, X has a probability distribution.
- How can we describe the distribution over an interval?
- We know the distribution is cannot be just a "list" of outcomes and their probabilities.

Continuous Random Variables vii

For discrete r.v., we have no worries.

Suppose Y has a **discrete Uniform distribution** with outcomes $S_Y = \text{even \#'s from 0 to 20} = \{0, 2, 4, \ldots, 18, 20\}$.

For any **discrete Uniform distribution** with N outcomes, each outcome is equally likely with probability $1/N$. Here, $N = 11$.

$$\begin{aligned}\sum_{\text{all } y} P(Y = y) = \sum_{\text{all } y} f(y) &= f(0) + f(2) + \cdots + f(18) + f(20) \\ &= (1/11) + (1/11) + \cdots + (1/11) + (1/11) \quad = \quad 1\end{aligned}$$

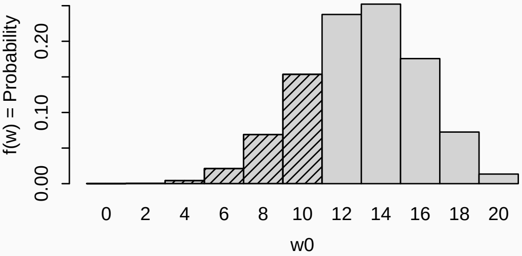
For a continuous r.v., we need a different definition of distribution.

cdf of Discrete Random Variables i

Recall the cumulative distribution function (cdf)

For any **discrete** r.v. W: $f(w)$ is a **probability mass func** (pmf)

$$\text{cdf } = F(w_0) = P(W \leq w_0) = \sum_{w \leq w_0} P(W = w) = \sum_{w \leq w_0} f(w)$$



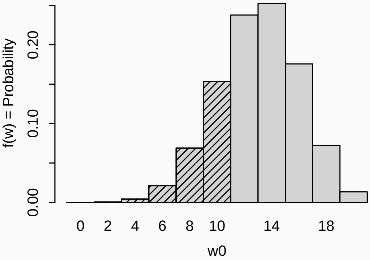
cdf of Discrete Random Variables ii

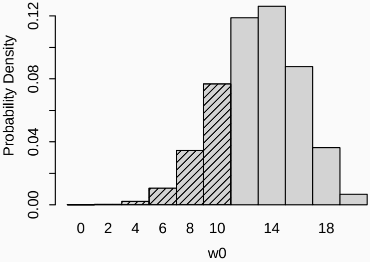
Probability density = probability per unit x.

The width of each bar = 2,

so height of each bar = probability density = probability / 2.

Area of each bar = probability. Combined area of all bars = 1



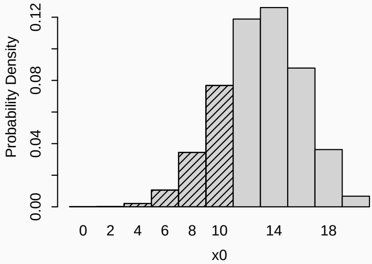


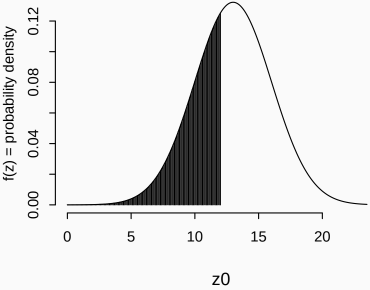
cdf is $F(w_0) = P(W \leq w_0)$ = sum of areas of all bars up to w_0.

cdf of Continuous Random Variables

$f(z)$ is a **probability density funcn** (pdf)

$$\text{cdf } = F(z_0) = P(Z \leq z_0) = \int_{z \leq z_0} \text{“}P(Z = z)\text{''} = \int_{z \leq z_0} f(z) dz$$





The cdf is $F(z_0) = P(Z \leq z_0) =$ sum of area under pdf up to z_0.
Total area under the pdf $= 1$.

Relationship between cdf and pdf for Continuous RV's

By the Fundamental Theorem of Calculus,

$$\text{pdf } = f(x) = \frac{d}{dx}\int_{-\infty}^{x} f(u)\, du = \frac{d}{dx}F(x) = F'(x)$$

So, you can get back and forth between the pdf $f(x)$ and cdf $F(x)$.

$$\text{cdf } = F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u) du$$

Uniform Distribution i

Now, back to our hard disk file location situation:

$X =$ the time it takes for a read/write head to locate a specific file header on a computer hard disk drive (once the head has been positioned over the correct track).

Since all 2ms intervals within [0,10) are equally likely, we say that X has a **continuous Uniform distribution**

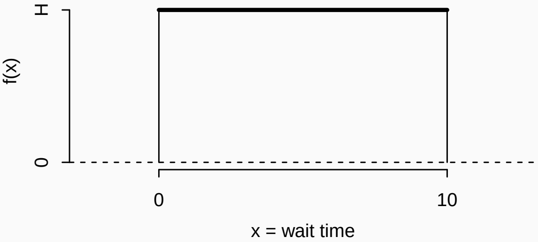
$P(2 \leq X \leq 4) = P(1 \leq X \leq 3) = P(5 \leq X \leq 7) =$ what value?

In fact, **all** intervals of equal length contained in [0,10) have the same probability.

Uniform Distribution ii

That is, the area under the curve is the same for **all** intervals equal length contained in [0,10).

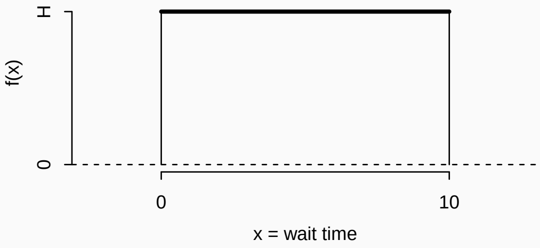
This observation allows us to propose a form for the probability density function, $f(x)$: a straight line.



What is the height, H?

Uniform Distribution iii

How tall is this pdf? That is, what is the value of H?



Total area under the curve above all possible x values must be $= 1$.

The width of the "box" is 10, with height H, and area $1 = 10(H)$.

Thus, $H = f(x) = 1/10 = 0.1$ for any x: $0 \leq x < 10$.

Uniform Distribution iv

Just to confirm the total area under the curve $= 1$...

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} f(x)\, dx = \int_{-\infty}^{\infty} H\, dx \\
&= 0 \;\; + \;\; \int_{0}^{10} H\, dx \;\; + \;\; 0 \\
&= Hx \Big|_{x=0}^{x=10} \\
&= 10 \times H - 0 \times H = 10 \times H
\end{aligned}$$

Therefore, $10 \times H = 1$,

which implies $H = 1/10 = 0.10$, as before.

Uniform Distribution v

OK. We had some questions at the start about the wait time X.

1. What is the chance of locating the file... in less than 2ms?

   $P(X \leq 2) = ?$

2. What is the chance of locating the file... in somewhere between 5ms and 7ms?

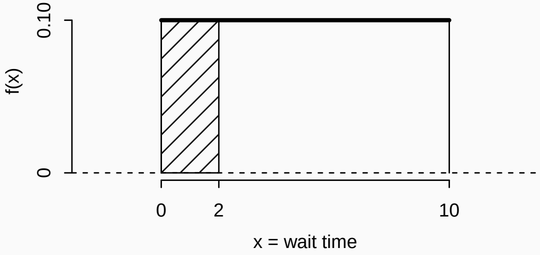
   $P(5 \leq X \leq 7) = ?$

3. What is the average wait time?
   (That is, the average time to locate the file.)

   $E(X) = ?$

Uniform Distribution vi

What is the chance of locating the file... in less than 2ms?



$$P(X \leq 2) = F(2) = \int_{-\infty}^{2} f(x)dx = \int_{0}^{2} (0.10)dx = 0.10(2-0) = 0.20$$

Uniform Distribution vi

What is $P(5 \leq X \leq 7) = ?$

General process to find probabilities of the form $P(a \leq X \leq b)$:

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b f(x)\,dx \\ &= \int_{-\infty}^b f(x)\,dx - \int_{-\infty}^a f(x)\,dx \\ &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) \end{aligned}$$

$P(5 \leq X \leq 7) = F(7) - F(5) = 7(0.10) - 5(0.10) = 0.20$

Continuous vs Discrete RV's

Comments and cautions about $F(b) - F(a)$:

For continuous random variables,

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

since $P(X = a) = \int_a^a f(x)\,dx = 0 \quad \text{and} \quad P(X = b) = 0.$

Be careful!

For discrete random variables,

when $P(Y = k) > 0$,

$$P(Y \le k) = F(k),$$

but... $\quad P(Y < k) = P(Y \le k - 1) = F(k - 1).$

What is the average wait time?

Since the uniform distribution is symmetric and the mean is always the balancing point of the distribution, we must have $E[X] = 5$, the midpoint of the range of X.

Expectation/Mean

For discrete random variables,

$$\mu = E(X) = \sum_{\text{all } x} x\, P(X = x) = \sum_{\text{all } x} x\, f(x)$$

For continuous random variables:

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

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_{0}^{10} \frac{x}{10}\, dx = \left.\frac{x^2}{20}\right|_0^{10} = \frac{10^2}{20} = \frac{100}{20} = 5$$

General Formula for Expectation of Uniform RV

If $X \sim \text{Uniform}(a, b)$, then you can show that

$$f(x) = \frac{1}{b-a} \text{ for } a \leq x \leq b$$

$$\mu = E(X) = \frac{a+b}{2} \qquad \text{(midpoint of the range of } X\text{)}$$

For X = wait time, $a = 0$ and $b = 10$, so

$$\mu = E(X) = \frac{a+b}{2} = \frac{0+10}{2} = \frac{10}{2} = 5$$

Variance for Continuous RV's

What about variance and standard deviation?

Variance (as always) = average squared distance from mean

$\sigma^2 = Var(X) = E[(X-\mu)^2]$

Simple guess for X = wait time standard deviation: $\sigma \approx 2.5$

For discrete random variables,

$$Var(X) = \sum_{\text{all } x} (x-\mu)^2 \; P(X=x) = \sum_{\text{all } x} (x-\mu)^2 \; f(x)$$

For continuous random variables:

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \; f(x) \; dx$$

Variance of Uniform RV

If $X \sim \text{Uniform}(a, b)$, then you can show that

$$\sigma^2 = E[(X - \mu)^2] = \frac{(b-a)^2}{12}$$

For $X =$ wait time, $a = 0$ and $b = 10$, so

$$\sigma^2 = Var(X) = \frac{(b-a)^2}{12} = \frac{(10-0)^2}{12} = \frac{100}{12} = 8.333$$

Standard deviation: $\sigma = SD(X) = \sqrt{8.333} = 2.887$