

Class 4: Continuous Random Variables

Alison Presmanes Hill
2017-10-5

Past: discrete rvs

Present: continuous rvs

- for discrete rvs, $P(X = x)$ is technically called the *probability mass function*
- for continuous rvs, the analogous concept is the *density*

Density

- The density of X can be seen as a value **proportional to** the chance of drawing from the population a number that is lying in the close proximity of X .
- Sadly, density does **not** give you probabilities directly
- Probabilities can only be obtained from densities by taking an **integral**
- Integrals are simply continuous **sums**

$f(x)$

For continuous X , $f(x)$ is a
Probability Density Function*

*Not a probability! (> 1)

Continuous rv pdf

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

$$2. f(x) \geq 0 \quad always$$

What does $P(X = a) = ?$

$$P(a < X < b) = \int_a^b f(x)dx$$

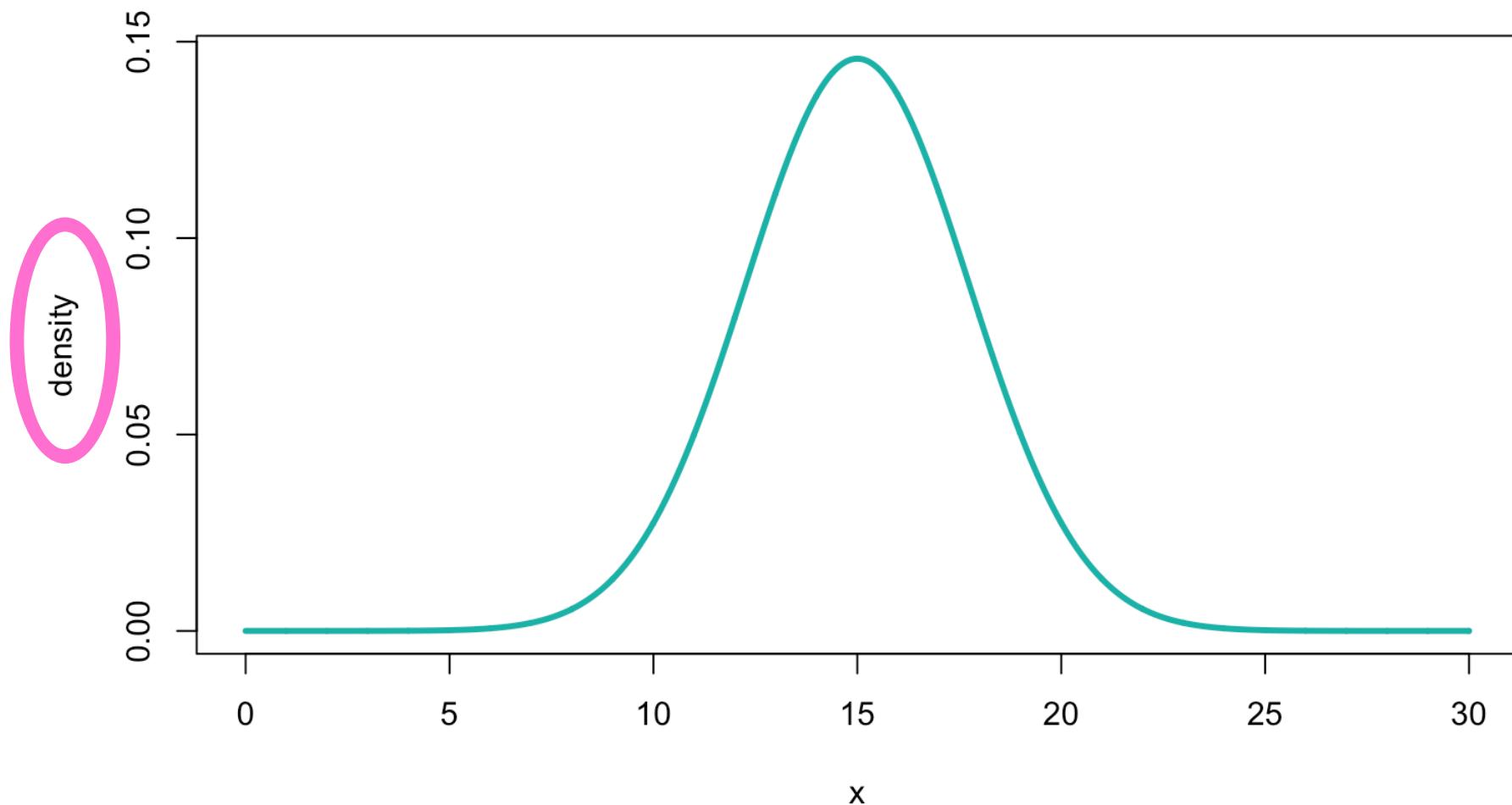
$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

One continuous distribution: Gaussian/normal

- Let's use the mean and sd from the coin flip example:
 - mean = 15
 - sd = 2.74

$$X \sim N(15, \sqrt{7.5})$$

**normal distribution pdf
mu = 15, var = 7.5**



Questions we can ask the pdf

- "What is the height of the density curve at $x = 10$?" 0.028

```
dnorm(10, 15, sqrt(7.5))
```



$$F(x)$$

For all x , $F(x)$ is a
cumulative distribution function

Cumulative distribution function

1. $F(x) \rightarrow 1$ as $x \rightarrow \infty$
2. $F(x) \rightarrow 0$ as $x \rightarrow -\infty$
3. $F(x)$ is monotonic; never decreasing
4. $F(x)$ does not need to be smooth
but is continuous

$$F(x) = P(X \leq x)$$

Discrete

$$F(x) = \sum_{t \leq x} f(t)$$

Continuous

$$F(x) = \int_{-\infty}^x f(t) dt$$

(since x is used as a variable in the upper limit of integration, we use some other variable, say “ t ”, in the integrand)

cdf in practice

- $F(x)$ is the probability of values less than x
- Thus, $F(x)$ is the probability of an interval
- If $F(x)$ is the cdf for the age in months of fish in a lake, then $F(10)$ is the probability a random fish is 10 months or younger.
- Can $F(10)$ be less than $F(9)$?

cdf for continuous rvs

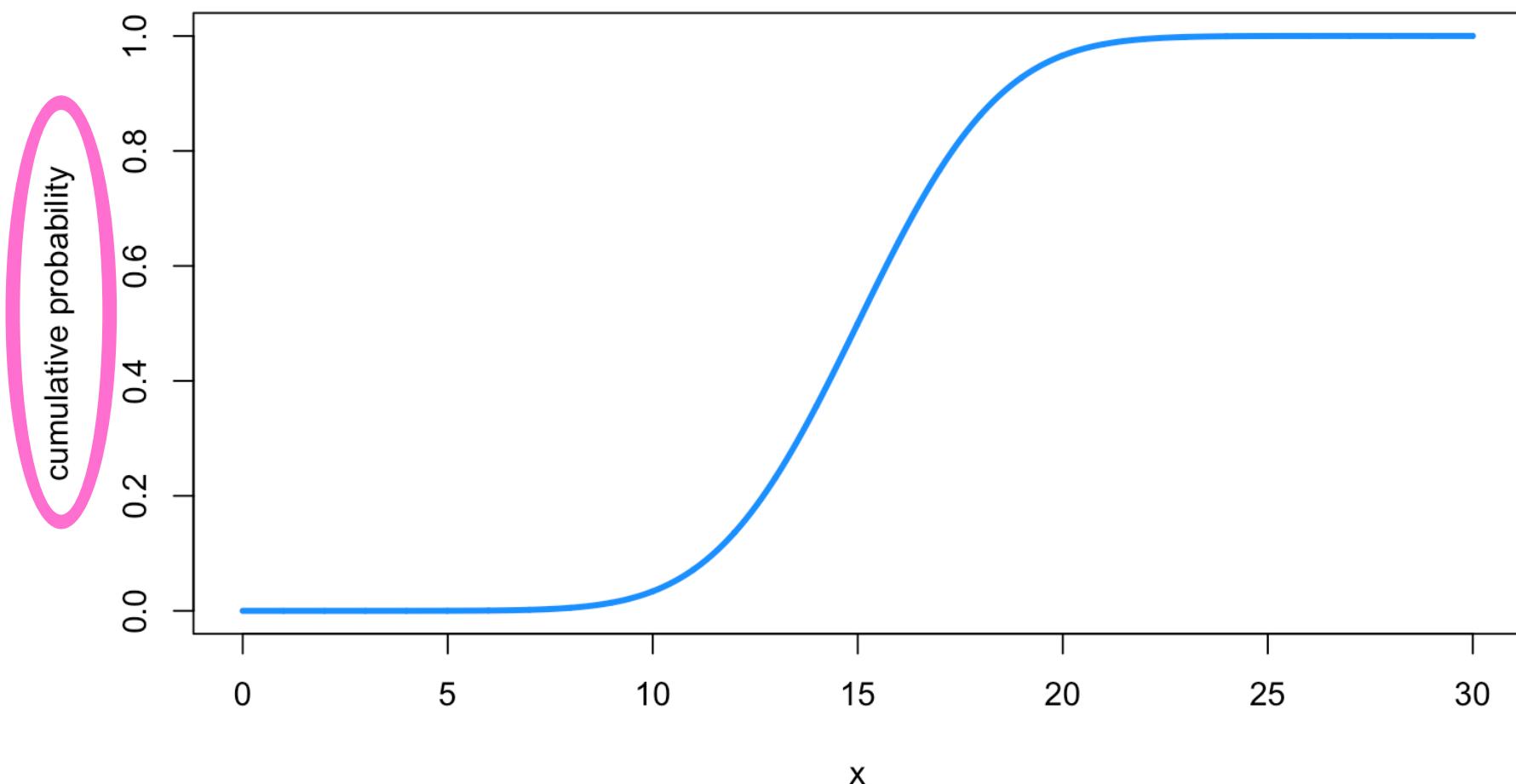
- By the Fundamental Theorem of Calculus:

$$F(b) - F(a) = \int_a^b f(x)dx$$

The area under the curve from a to b of a function f is just the difference between the values of that function's antiderivative, F , at b and a .

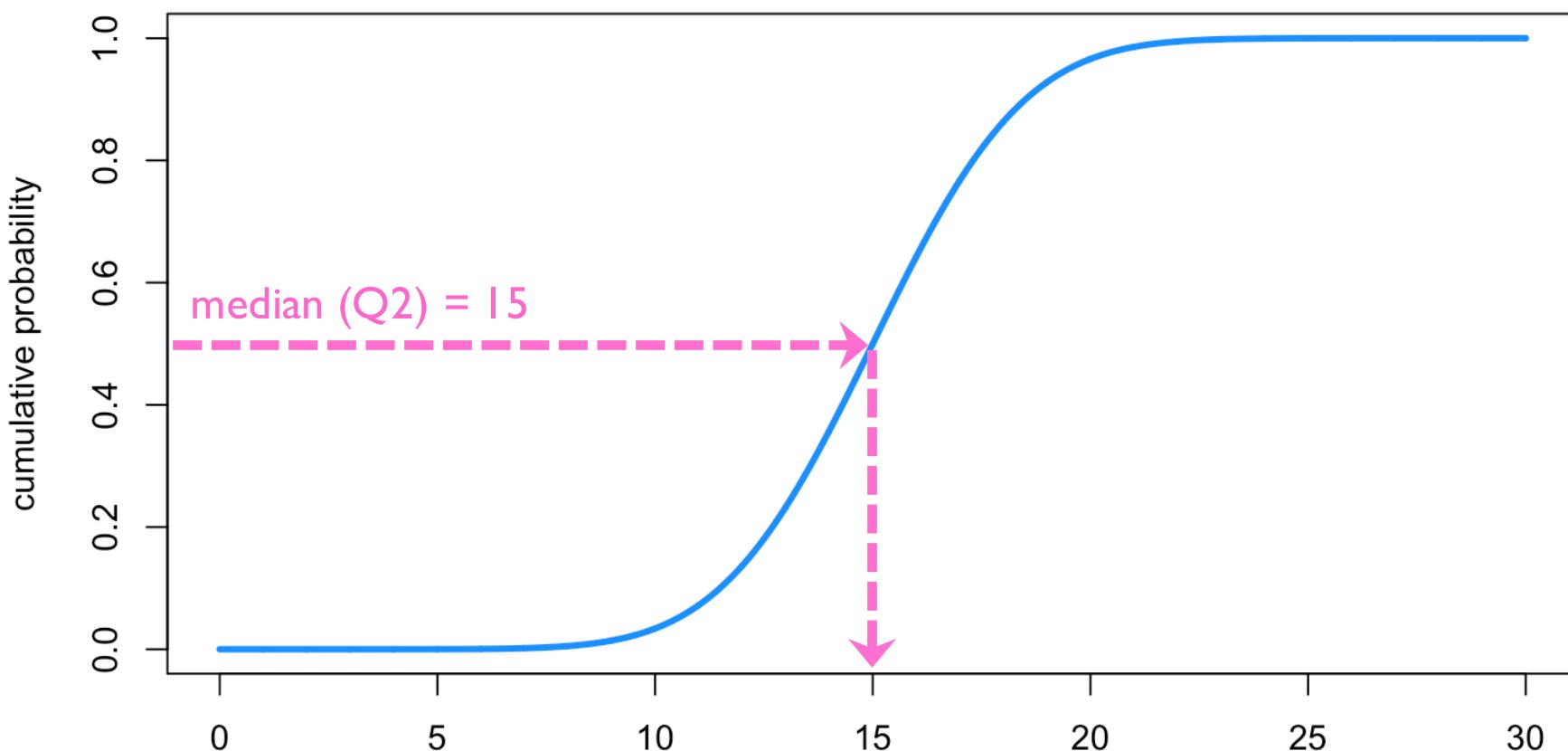
$$X \sim N(15, \sqrt{7.5})$$

normal distribution cdf
mu = 15, var = 7.5



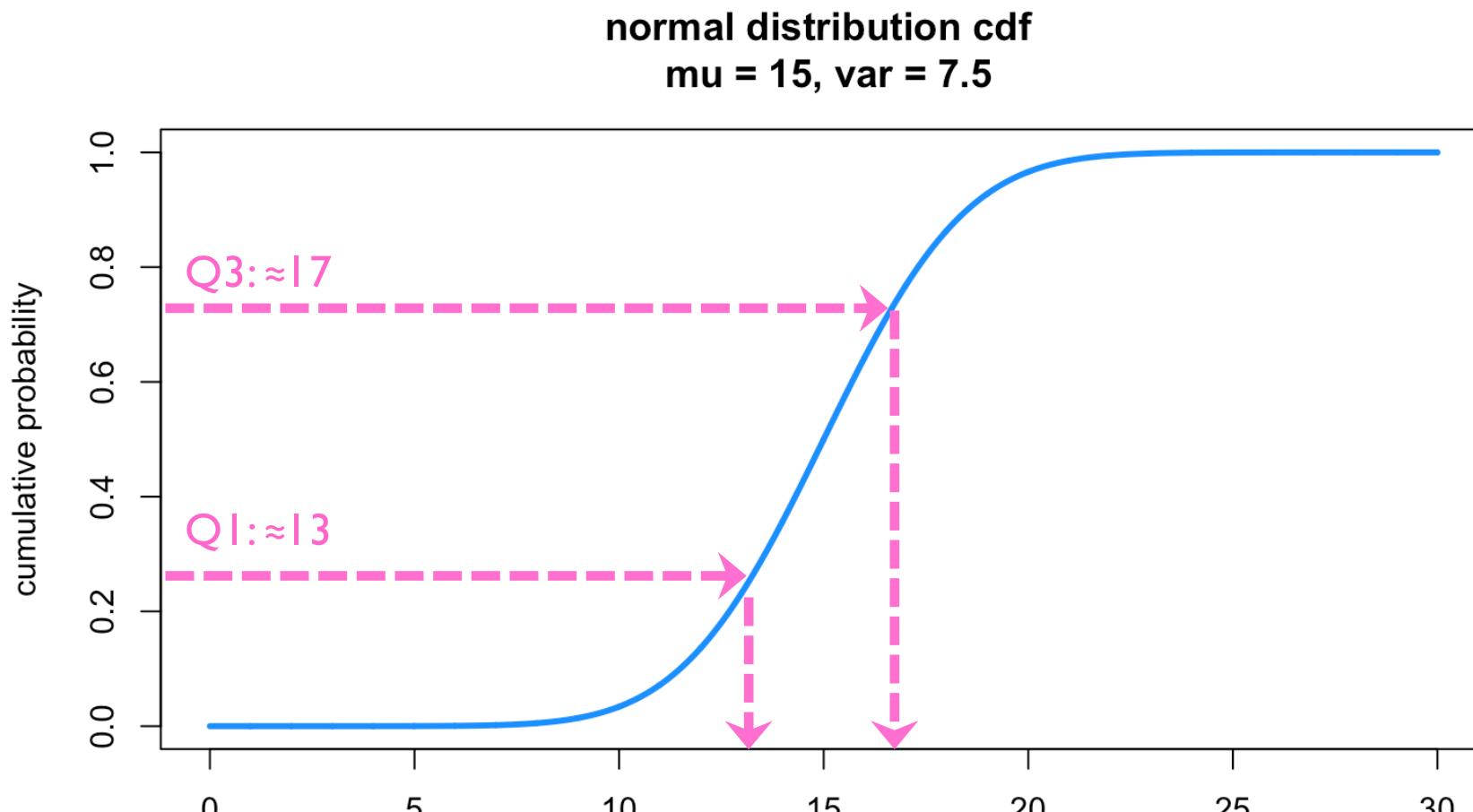
$$X \sim N(15, \sqrt{7.5})$$

normal distribution cdf
mu = 15, var = 7.5



```
> qnorm(.5, 15, sqrt(7.5)) #Q2 (median)
[1] 15
```

$$X \sim N(15, \sqrt{7.5})$$



```
x > qnorm(.25, 15, sqrt(7.5)) #Q1
[1] 13.15283
> qnorm(.75, 15, sqrt(7.5)) #Q3
[1] 16.84717
```

Questions we can ask the cdf in R

- "What is the probability that x is exactly 7?" 0
- "What is the probability that x is between 18 to 24?" 0.136
- "What is the probability that x ends in 7?" 0
- "What is the probability that x is greater than 24?" 0.0005

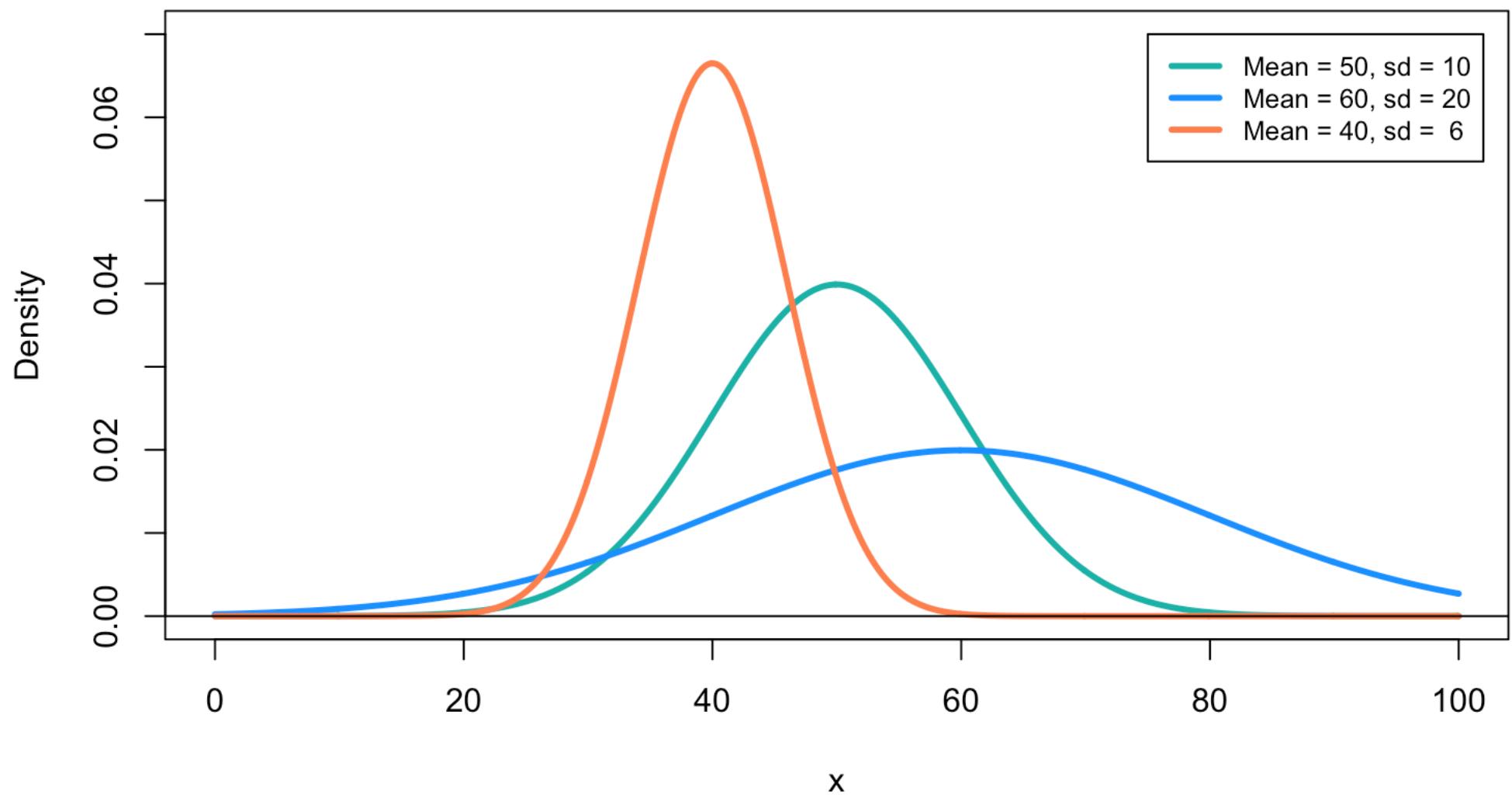


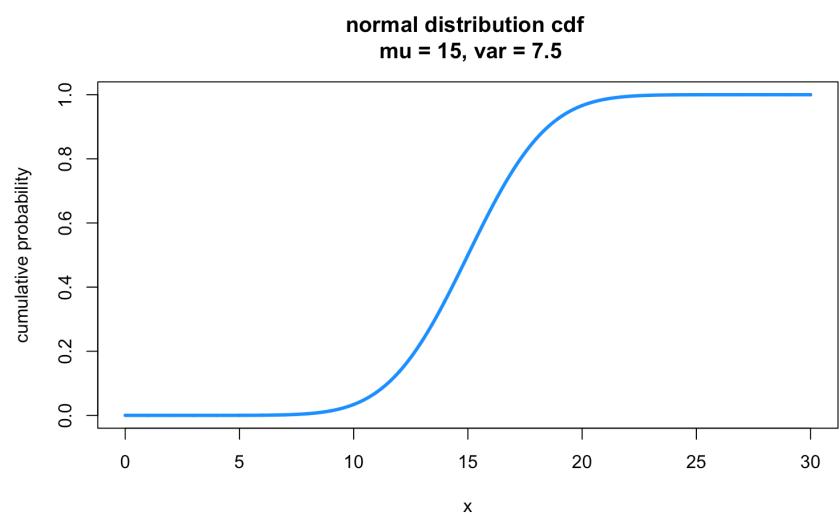
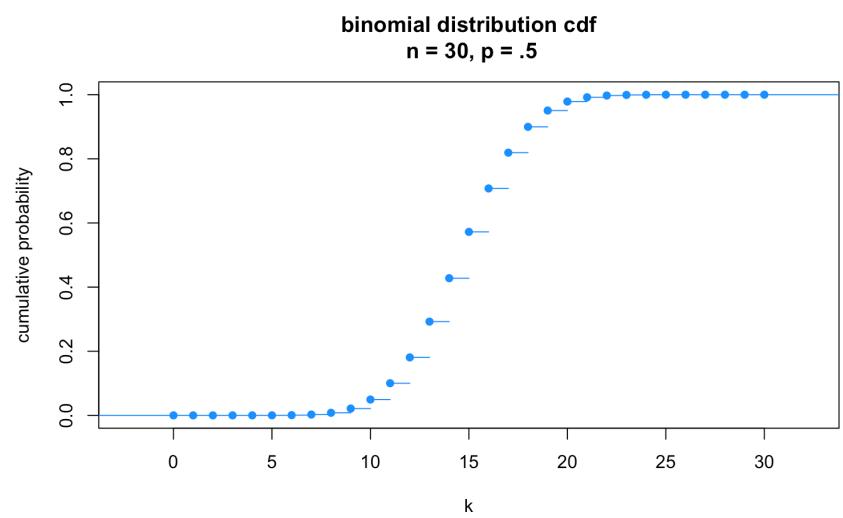
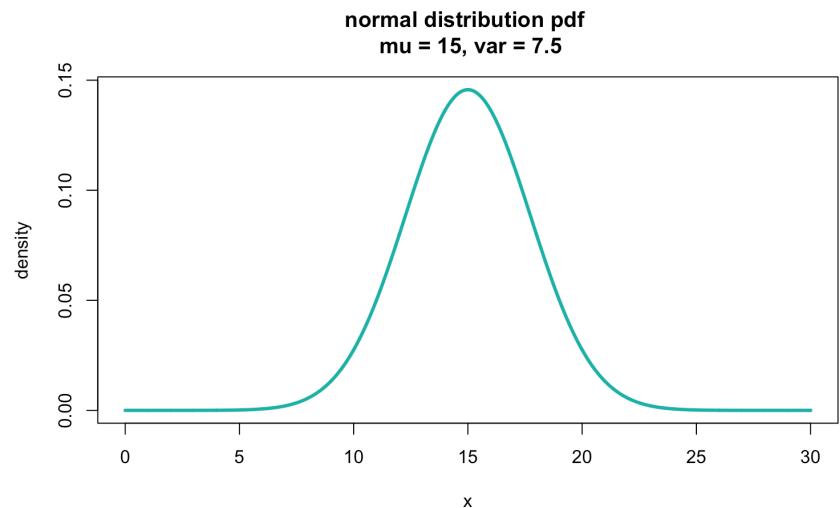
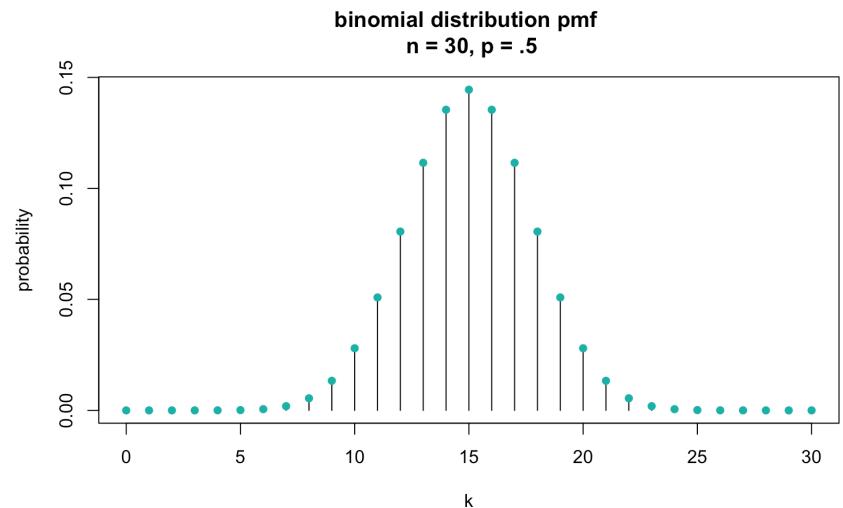
Questions we can ask the cdf in R

- "What is the probability that x is exactly 7?" 0
- "What is the probability that x is between 18 to 24?" 0.136
`pnorm(24, 15, sqrt(7.5)) - pnorm(18, 15, sqrt(7.5))`
- "What is the probability that x ends in 7?" 0
- "What is the probability that x is greater than 24?" 0.0005
`pnorm(24, 15, sqrt(7.5), lower.tail = FALSE)`



Normal Distributions





Transformations

- If $X \sim N(\mu, \sigma^2)$ and $Y = X + b$
 - Then $Y \sim N(\mu + b, \sigma^2)$
- If $X \sim N(\mu, \sigma^2)$ and $Y = aX$
 - Then $Y \sim N(a\mu, a^2\sigma^2)$
- If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$
 - Then $Y \sim N(a\mu + b, a^2\sigma^2)$

Random variables

Discrete

- Finite or countably infinite sample space.
- Subset of integers.
- Use sums.
- Has **pmf**.

Continuous

- Uncountably infinite sample space.
- Subset of the real number line.
- Use integrals.
- Has **pdf**.

continuous rvs

Outside support, $f(x)=0$

$$f(x)$$

Integrate
from $-\infty$ to x

$$F(x)$$

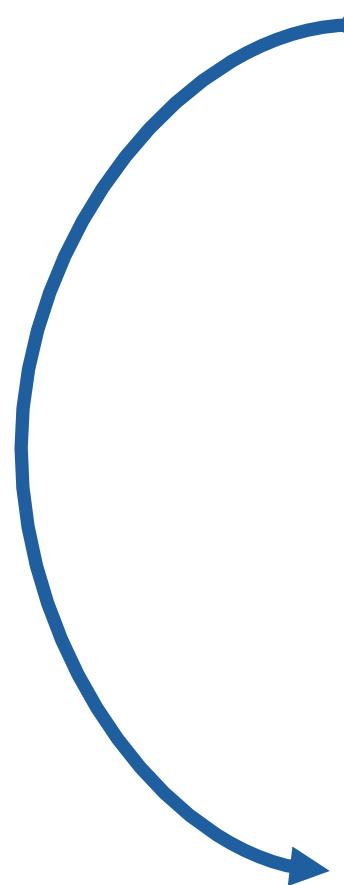
Differentiate

Outside support,
 $F(x)=0$ or $F(x)=1$

X is a continuous rv

$$f(x) = \begin{cases} 4x & : 0 \leq x \leq 1/2 \\ -4x + 4 & : 1/2 \leq x \leq 1 \end{cases}$$

Integrate
from $-\infty$ to x



$$F(x) = \begin{cases} 2x^2 + A & : 0 \leq x \leq 1/2 \\ -2x^2 + 4x + C & : 1/2 \leq x \leq 1 \end{cases}$$

Wolfram Alpha:

integrate $4x = 2x^2 + \text{constant}$

integrate $-4x + 4 = -2x^2 + 4x + \text{constant}$

How to solve for A and C

$$F(x) = \begin{cases} 2x^2 + A & : 0 \leq x \leq 1/2 \\ -2x^2 + 4x + C & : 1/2 \leq x \leq 1 \end{cases}$$

Where A and C are constants of integration. Now we use the properties of $F(x)$ to solve for those constants.

Specifically:

1. We know: $F(x_{min}) = 0$; so $F(0) = 0$
2. We also know: $F(x_{max}) = 1$; so $F(1) = 1$

Solve for A first

$$F(x) = \begin{cases} 2x^2 + A & : 0 \leq x \leq 1/2 \\ -2x^2 + 4x + C & : 1/2 \leq x \leq 1 \end{cases}$$

$$F(x = 0) = 0 = 2(0^2) + A, \text{ so } A = 0$$

Solve for C next

$$F(x) = \begin{cases} 2x^2 + A & : 0 \leq x \leq 1/2 \\ -2x^2 + 4x + C & : 1/2 \leq x \leq 1 \end{cases}$$

$$F(x = 1) = 1 = -2(1^2) + (4 \times 1) + C, \text{ so } C = -1$$

Plug in: $A = 0$; $C = -1$

$$F(x) = \begin{cases} 2x^2 + A & : 0 \leq x \leq 1/2 \\ -2x^2 + 4x + C & : 1/2 \leq x \leq 1 \end{cases}$$

$$F(x) = \begin{cases} 2x^2 & : 0 \leq x \leq 1/2 \\ -2x^2 + 4x - 1 & : 1/2 \leq x \leq 1 \end{cases}$$

Alternative way to solve for C

- $F(x)$ must be continuous, meaning that the graph of $F(x)$ must touch when $x = 1/2$ (the two pieces of the piecewise graph must touch). So, to solve for C , we set $x = 1/2$ for each equation, plug in $A = 0$, and make them equal each other.

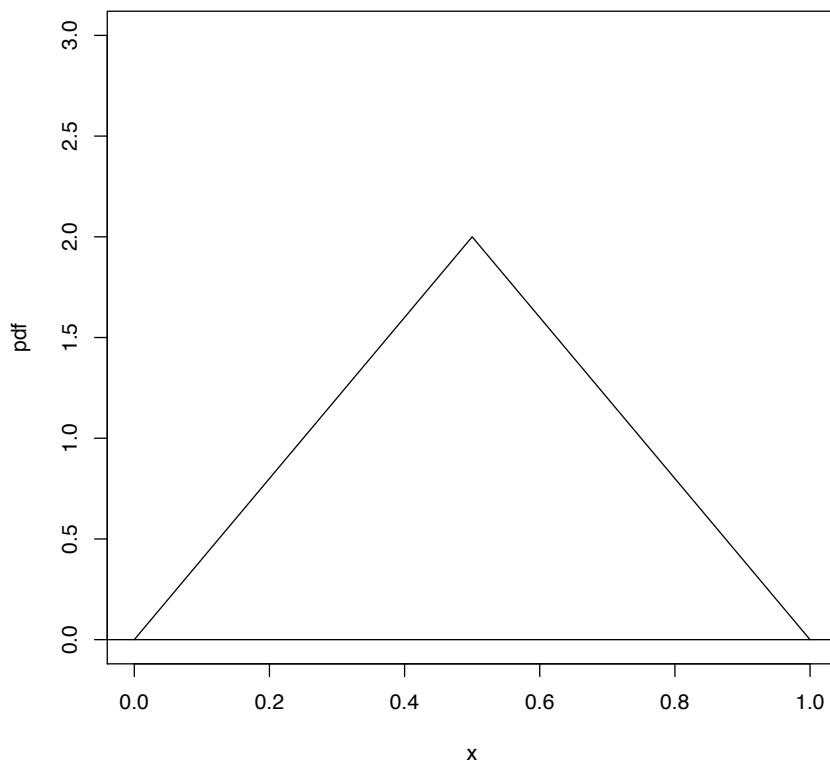
$$F(.5) = 2(.5^2) = -2(.5^2) + 4(.5) + C$$

$$F(.5) = .5 = 1.5 + C$$

$$C = .5 - 1.5 = -1$$

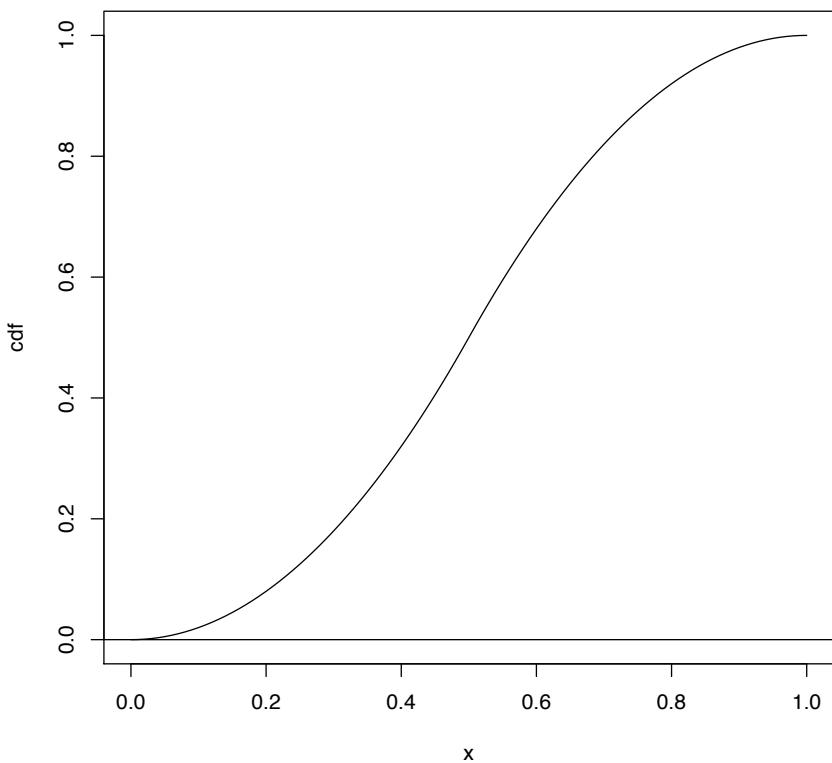
$$f(x) = \begin{cases} 4x & : 0 \leq x \leq 1/2 \\ -4x + 4 & : 1/2 \leq x \leq 1 \end{cases}$$

pdf



$$F(x) = \begin{cases} 2x^2 & : 0 \leq x \leq 1/2 \\ -2x^2 + 4x - 1 & : 1/2 \leq x \leq 1 \end{cases}$$

cdf



continuous rvs

Outside support, $f(x)=0$

$$f(x)$$

Integrate
from $-\infty$ to x

$$F(x)$$

Differentiate

Outside support,
 $F(x)=0$ or $F(x)=1$

X is a continuous rv

$$f_X(x) = \begin{cases} 1/2 & -1 \leq x < 1, \\ 0 & \text{otherwise} \end{cases}$$

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

Differentiate

Wolfram Alpha:
differentiate 0 = 0
differentiate $(x+1)/2 = 1/2$
differentiate 1 = 0



Expectation of a
Continuous random variable

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

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

Population median from pdf

- Halfway point: half the population has a lower value, and half has a higher value
- If $f(x)$ is a pdf, then the median of the distribution is the point M such that:

$$\int_{-\infty}^M f(x)dx = 0.5$$

Example: ladybug median life span

- Suppose a ladybug's life span (in months) has a pdf:

$$f(x) = \begin{cases} \frac{1}{72}x, & 0 \leq x \leq 12 \\ 0, & \text{otherwise} \end{cases}$$

- What is the **median** life span for the ladybug population?



Example: ladybug median life span

$$\frac{1}{2} = \int_0^M \frac{1}{72} x dx = \frac{1}{144} x^2 \Big|_0^M = \frac{1}{144} M^2$$

$$\frac{1}{2} = \frac{1}{144} M^2$$

$$72 = M^2$$

$$M = \sqrt{72} = 8.49$$



Aside: wolfram alpha

Wolfram|Alpha Step-by-step Solution

 **Wolfra** Definite integral:

integrate $x dx$ from 0 to m

Compute the definite integral:
$$\int_0^m x dx$$

Definite integral:

$$\int_0^m x dx = \frac{m^2}{2}$$

Apply the fundamental theorem of calculus.
The antiderivative of x is $\frac{x^2}{2}$:

$$= \frac{x^2}{2} \Big|_0^m$$

$$f(8.49) = \int_0^{8.49} \frac{1}{72} x dx = 0.5$$

- If we had the cdf, we could check this result:

$$F(M) = F(8.49) = 0.5$$

- How do we get there from the pdf? **Integrate!**

I know what you are thinking- I do so wish we could do more integration!



$$f(x) = \begin{cases} \frac{1}{72}x, & 0 \leq x \leq 12 \\ 0, & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} \frac{1}{144}x^2 + \text{constant}, & 0 \leq x \leq 12 \\ 0, & \text{otherwise} \end{cases}$$

constant = 0 (*do this one on your own*)

$$F(x) = \begin{cases} \frac{1}{144}x^2, & 0 \leq x \leq 12 \\ 0, & \text{otherwise} \end{cases}$$

```
> (1/144) * (8.49^2)
[1] 0.5005563
```

$$F(M) = F(8.49) = 0.5$$

pdf → cdf:
Integrate!






Using the cdf

- Now we can answer so many more questions without having to integrate each time...
- Q1? 6
- Q3? 10.4
- IQR? 4.4



Example: ladybug mean life span

- Suppose a ladybug's life span has a pdf:

$$f(x) = \begin{cases} \frac{1}{72}x, & 0 \leq x \leq 12 \\ 0, & \text{otherwise} \end{cases}$$

- What is the **mean** life span for the ladybug population?

$$\int_{-\infty}^{+\infty} x f(x) dx$$



Example: ladybug mean life span

$$\int_{-\infty}^{+\infty} x f(x) dx = \int_0^{12} x \left(\frac{1}{72}x\right) dx$$

$$= \int_0^{12} \frac{1}{72} x^2 dx$$

$$= \frac{1}{216} x^3 \Big|_0^{12} = 8$$



Example: ladybug life spans

- What is the **median** life span? **8.49 months**
- What is the **mean** life span? **8 months**
- The mean is < median, so more than 50% of ladybugs live longer than 8 months
- What is the **variance** of ladybug life spans



I'll leave this for you to solve...

Variance of ladybug life spans

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

12



$$\int_{-\infty}^{\infty}$$

8



$$f(x) = \begin{cases} \frac{1}{72}x, & 0 \leq x \leq 12 \\ 0, & \text{otherwise} \end{cases}$$

0



Mini-review: probability functions

Discrete rvs:
probability mass function (pmf)

- $f(x)$ is defined by the distribution!
- Famous ones:
 - Bernoulli
 - Binomial
 - Negative binomial
 - Geometric
 - Hypogeometric
 - Poisson

Continuous rvs:
probability density function (pdf)

- $f(x)$ is defined by the distribution!
- Famous ones:
 - Normal/Gaussian
 - Chi-squared
 - F
 - Student's t
 - Gamma
 - Beta

Mini-review: conditions for probability functions

Probability	Discrete rv: $f(x)$ is a pmf if...	Continuous rv: $f(x)$ is a pdf if...
$P(A) \geq 0$ for all $A \in \Omega$	$f(x) > 0$ for all $x \in \Omega$	$f(x) > 0$ for all $x \in \mathbb{R}$
$P(\Omega) = 1$	$\sum_{x_i \in \Omega} f(x_i) = 1$	$\int_{\mathbb{R}} f(x) = 1$

Mini-review: cumulative density functions

Discrete rvs

$$F(x) = \sum_{t \leq x} f(t)$$

Continuous rvs

$$F(x) = \int_{-\infty}^x f(t) dt$$

where $f(t)$ is just the pmf/pdf

Mini-review: Expectation of an rv

Discrete rvs

$$E(X) = \sum_{\text{all } x} xf(x)$$

Continuous rvs

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Mini-review: Variance of an rv

Discrete rvs

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

Continuous rvs

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

What you need to know about distributions

- That lots of distributions exist (not just the normal!), and what types of rvs can typically be represented by them
- What the pmf/pdf represents in terms of probability
- What the cdf represents in terms of probability
- How to calculate the expectation value and variance of the distribution, given the pmf/pdf
- How to determine the pdf from the cdf for any function
 - Requires differentiation
- How to determine the cdf from the pdf for any function
 - Requires integration