Stat310

CDF & Normal distribution

Hadley Wickham

1. Test information
2. Word problem practice
3. CDF
4. Normal distribution

Test

Important dates

Feb 9: last class covered in test

Feb 14: in class review

Feb 15: test available

Feb 16: homework 5 due

Feb 20: test due (4pm)

Mar 8: Stats in practice essay due

Mar 8: Homework 6 due

Test

120 minute take home test. 4 questions.

Covers everything up to Feb 9: probability, discrete random variables, continuous random variables. See website.

Approximately half applied (working with real problems) and half theoretical (working with mathematical symbols).

Honour code

No collaboration.

No communication about the questions or your answers.

Only outside resources allowed are: a one-page double-sided note sheet and wolfram alpha.

Pledged and signed.

Expectations

Points will be awarded for fully converting a word problem into a mathematical problem.

You should be able to recall any fact on the basic math sheet.

I will supply random mathematical facts and tables of probabilities (if needed). You may use wolframalpha, but it will not be necessary.

Note sheet

Much of the usefulness of a note sheet is the process of making it.

You want to condense everything we have covered. Pull out ongoing themes. Make tables. Use colour!

Not useful: a photocopy of someone else's notes, a verbatim copy of the textbook.

Examples online.

Word problems

Practice

Last season, Ron Artest of the Houston Rockets made 141 three pointers out of 347 attempts.

How many 3 pointers do you expect him to make before he misses one? What's the probability he makes the next 5 shots then misses the 6th?

What assumptions did you make?

Your turn

What event are we interested in?

What's the distribution of this event?

What are the questions asking in statistical terms?

Practice

On average, Ron Artest makes a 3pt attempt every 190 seconds.

How long do you expect to wait for him to make 6 attempts? What's the probability you wait more than 5 minutes before his next 3pt attempt?

How many attempts do you expect him to make in 10 minutes? What's the probability he makes one or less?

Your turn

What event are we interested in?

What's the distribution of this event?

What are the questions asking in statistical terms?

CDF

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

discrete

continuous

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

discrete

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

continuous

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

discrete

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

continuous

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

$$\lim_{x \to -\infty} F(x) = \text{[redacted]}$$

$$\lim_{x \to \infty} F(x) = \text{[redacted]}$$

monotone [redacted]

[redacted]-continuous

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to \infty} F(x) = 1$$

monotone increasing

right-continuous

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to \infty} F(x) = 1$$

monotone increasing

right-continuous

Conversely, if a function has these properties, then it's a cdf

Using the cdf

$$P(a < X \leq b) = F(b) - F(a)$$

Exact computation

Tables

Computer

Exact computation

For some distributions we can write the cdf in closed form. For example: the exponential distribution has cdf:

$$F(x) = \begin{cases} 1 - e^{-x/\theta}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Exponential pdf

$$f(x) = \frac{1}{\theta} e^{-x/\theta}$$

$$F(x) = \int_0^x \frac{1}{\theta} e^{-t/\theta} dt \quad x > 0$$

```
integrate 1/theta exp(-t /theta) dt from 0 to x
```

Memorylessness

$X \sim \text{Exponential}(\theta)$

$P(X > x + y \mid X > x)$?

What's the probability we wait y more minutes, given that we've already waited x?

F(x) usually has **three** parts:

$x < \mathit{min}$

$\mathit{min} < x < \mathit{max}$

$x > \mathit{max}$

F(x) usually has **three** parts:

x < *min*

min < x < *max*

x > *max*

$$F(x) = \begin{cases} 1 - e^{-x/\theta}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Your turn

On average, Ron Artest makes a 3pt attempt every 190 seconds. What's the probability you wait more than 5 minutes before his next 3pt attempt?

Exact computation

For many distributions we can't write the cdf in closed form:

http://en.wikipedia.org/wiki/Binomial_distribution

http://en.wikipedia.org/wiki/Gamma_distribution

And similarly for the next distribution we'll learn about: the normal distribution

Normal distribution

Parameters and pdf

The normal distribution has two parameters:

$\mu \in R$, the mean

$\sigma > 0$, the standard deviation
(or $\sigma^2 > 0$, the variance)

And has pdf...

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Is this a valid pdf?

Note that this pdf is symmetric about mean!

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Is this a valid pdf?

```
integrate 1/(sigma sqrt(2 pi)) e ^ (-(x- mu)^2 /
(2(sigma^2))) dx from -infinity to infinity
```

Note that this pdf is symmetric about mean!

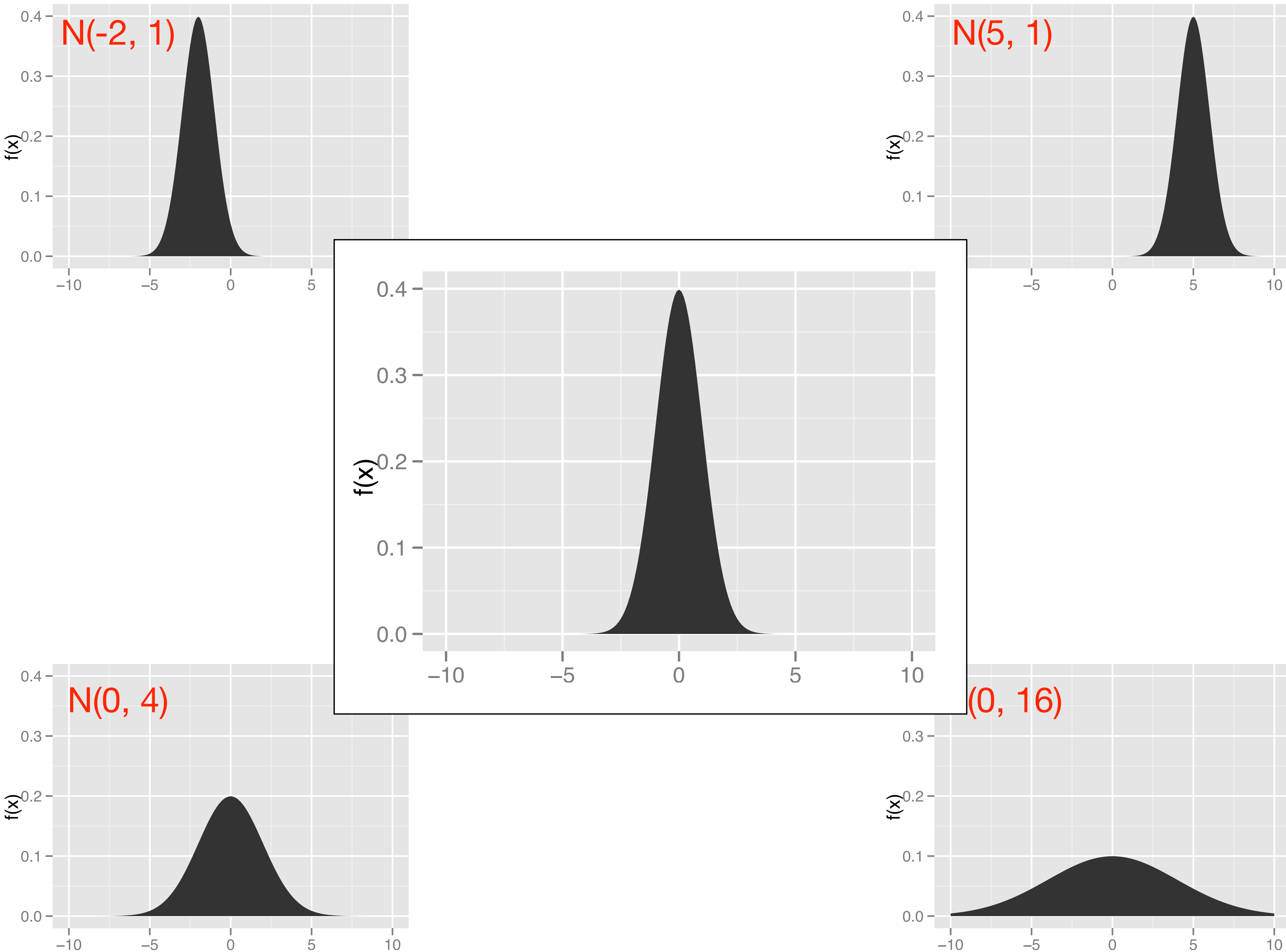
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Is this a valid pdf?

```
integrate 1/(sigma sqrt(2 pi)) e ^ (-(x- mu)^2 /
(2(sigma^2))) dx from -infinity to infinity
```

```
integrate 1/(sqrt(2 pi)) e ^ (-x^2 / 2) dx from -
infinity to infinity
```

Note that this pdf is symmetric about mean!


N(-2, 1)
N(5, 1)
N(0, 4)
(0, 16)
f(x)
0.0
0.1
0.2
0.3
0.4
−10
−5
0
5
10

When to use

Empirically arises in many real situations, typically when a measurement can be thought of as the sum or mean of other measurements.

In a few weeks we’ll see why this true theoretically (the CLT)

Moments

$$M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

What is the mean and variance of the normal distribution?

With wolfram alpha

```
d/dt e^(mu*t + 1/2 sigma^2 t^2) at t = 0

d^2/dt^2 e^(mu*t + 1/2 sigma^2 t^2) at t
= 0

d^2/dz^2 exp(mu*z + 1/2 sigma^2 z^2) at
z = 0
```

Transformations

If $X \sim \text{Normal}(\mu, \sigma^2)$, and $Y = aX + b$

THEN $Y \sim \text{Normal}(a\mu + b, a^2\sigma^2)$

Standard normal

If $b = -\mu$ and $a = 1/\sigma$, we often write

$Z = (X - \mu) / \sigma$

$Z \sim \text{Normal}(0, 1)$ = **standard normal**

We'll see why this is so important shortly

Your turn

Assume the weight of a bag of m&m's is normally distributed with a mean of 50 grams with a variance of 4 grams.

What's the probability that you get a bag that weighs more than 55 grams? (How would you work it out?)

Normal cdf

$$P(Z < z) = \Phi(z)$$

So commonly used that it has it's own symbol.

But what is it?

```
integrate 1/(sigma sqrt(2 pi)) e ^ (-(t -
mu)^2 / (2(sigma^2))) dt from -infinity to
x
integrate 1/(sqrt(2 pi)) e ^ (-t^2 / 2) dt
from -infinity to x
```

Closed form

Neither the CDF of the normal distribution nor erf can be expressed in terms of finite additions, subtractions, multiplications, and root extractions, and so both must be either computed numerically or otherwise approximated.

i.e. can't be expressed in **closed form**

http://mathworld.wolfram.com/NormalDistribution.html

Tables

Two approaches

If you don’t have a closed form solution you can:

Look up the number in a table.

Problem: you need a table for every combination of the parameters

Use a computer.

Problem: you need a computer

Standard normal

Fortunately, for the normal distribution, we can convert any random variable with a normal distribution to a **standard normal**.

This means we only need one table for any possible normal distribution. (For other distributions there will be multiple tables, and typically you will have to pick one with similar values to your example).

Using the tables

Column + row = z

Find: $\Phi(2.94)$, $\Phi(-1)$, $\Phi(0.01)$, $\Phi(4)$

Tricks: $\Phi(0) = 0.5$, $\Phi(-x) = 1 - \Phi(x)$

Can also use in reverse: For what value of z is $P(Z < z) = 0.90$? i.e. What is $\Phi^{-1}(0.90)$?

Find: $\Phi^{-1}(0.1)$, $\Phi^{-1}(0.5)$, $\Phi^{-1}(0.65)$, $\Phi^{-1}(1)$

Your turn

Assume the weight of a bag of m&m's is 50 grams with a variance of 4 grams.

What's the probability that you get a bag that weighs more than 55 grams?

(Convert to standard normal and then look up in table)