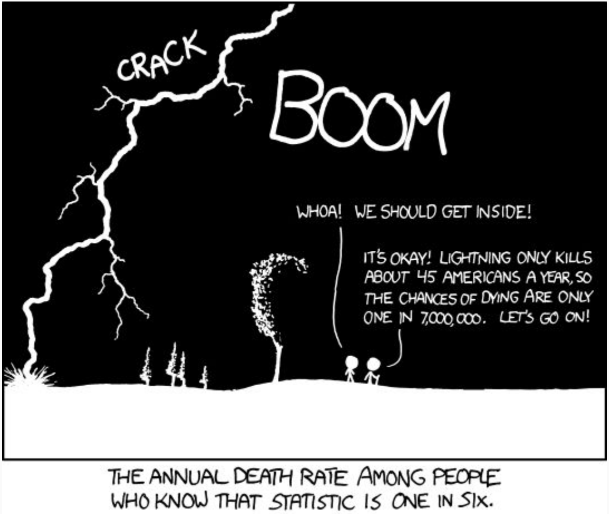
Political Science 209 - Fall 2018

Probability II

Florian Hollenbach

7th November 2018

Conditional Probability



Conditional Probability

Sometimes information about one event can help inform us about likelihood of another event

Examples?

Conditional Probability

Sometimes information about one event can help inform us about likelihood of another event

Examples?

- What is the probability of rolling a 5 and then a 6?
- What is the probability of rolling a 5 and then a 6 given that we rolled a 5 first?

Conditional Probability

If it is cloudy outside, gives us additional information about likelihood of rain

If we know that one party will win the House, makes it more likely that party will win certain Senate races

Independence

If the occurrence of one event (A) gives us information about likelihood of another event, then the two events are not independent.

Independence

If the occurrence of one event (A) gives us information about likelihood of another event, then the two events are not independent.

Independence of two events implies that information about one event does not help us in knowing whether the second event will occur.

Independence

For many real world examples, independence does not hold

Knowledge about other events allows us to improve guesses/probability calculations

Independence

When two events are independence, the probability of both happening is equal to the individual probabilities multiplied together

$P(A \mid B)$

Probability of A given/conditional that B has happened

Conditional Probability

$\mathrm{P(A \mid B)} = \frac{P(AandB)}{P(B)}$

Probability of A and B happening (joint) divided by probability of B happening (marginal)

Conditional Probability

Definitions:

P(A and B) - joint probability

P(A) - marginal probability

Conditional Probability

$\text{P}(\text{rolled 5 then 6}) = ?$

Conditional Probability

P(rolled 5 then 6) = ?

P(rolled 5 then 6) $= \frac{1}{36}$

P(rolled 5 then 6 | 5 first) $= \frac{P(5 then 6)}{P(5)}$

Conditional Probability

P(rolled 5 then 6) = ?

P(rolled 5 then 6) $= \frac{1}{36}$

P(rolled 5 then 6 | 5 first) $= \frac{P(5 then 6)}{P(5)}$

$\frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$

Conditional Probability

The probability that it is Friday and that a student is absent is 0.03. What is the probability that student is absent, given that it is Friday?

$P(\text{absent} \mid \text{Friday}) = ?$

Conditional Probability

The probability that it is Friday and that a student is absent is 0.03. What is the probability that student is absent, given that it is Friday?

$\mathrm{P}(\text{absent} \mid \text{Friday}) = ?$

$\mathrm{P}(\text{absent} \mid \text{Friday}) = \frac{0.03}{0.2} = 0.15$

Conditional Probability

$\text{P(A} \mid \text{B)} = \frac{P(AandB)}{P(B)}$

Also means:

$\text{P(A and B)} = \text{P(A} \mid \text{B) P(B)}$

Independence

If A and B are independent, then

- $P(A \mid B) = P(A)$ & $P(B \mid A) = P(B)$
- $P(A \text{ and } B) = P(A) \times P(B)$

Independence

If A|C and B|C are independent, then

- P(A and B | C) = P(A |C) $\times$ P(B | C)

What is the probability of drawing any card between 2 and 10, or jack, queen, king in any color?

Probability Problems

What is the probability of drawing two kings from a full deck of cards?

Probability Problems

What is the probability of drawing two kings from a full deck of cards?

$P(2 \text{ kings}) = \frac{4}{52} \times ?$

Probability Problems

What is the probability of drawing two kings from a full deck of cards?

$P(2 \text{ kings}) = \frac{4}{52} \times ?$

$P(2 \text{ kings}) = \frac{4}{52} \times \frac{3}{51} = \frac{12}{2652} = \frac{1}{221}$

Probability Problems

Annual income	Took 209	Took 309	TOTAL
Under \$50,000	36	24	60
\$50,000 to \$100,000	109	56	165
over \$100,000	35	40	75
Total	180	120	300

Is the probability of making over \$100,000 and the probability of having taken 309 independent?

Probability Problems

Annual income	Took 209	Took 309	TOTAL
Under \$50,000	36	24	60
\$50,000 to \$100,000	109	56	165
over \$100,000	35	40	75
Total	180	120	300

Is the probability of making over \$100,000 and the probability of having taken 309 independent?

P(over \$100k & 309) = P(over \$100k) $\times$ P(309)?

Probability Problems

Annual income	Took 209	Took 309	TOTAL
Under \$50,000	36	24	60
\$50,000 to \$100,000	109	56	165
over \$100,000	35	40	75
Total	180	120	300

What is the probability of any student making over \$100,000?

Probability Problems

Annual income	Took 209	Took 309	TOTAL
Under \$50,000	36	24	60
\$50,000 to \$100,000	109	56	165
over \$100,000	35	40	75
Total	180	120	300

What is the probability of a student making over \$100,000, conditional that he took 309?

Probability Problems

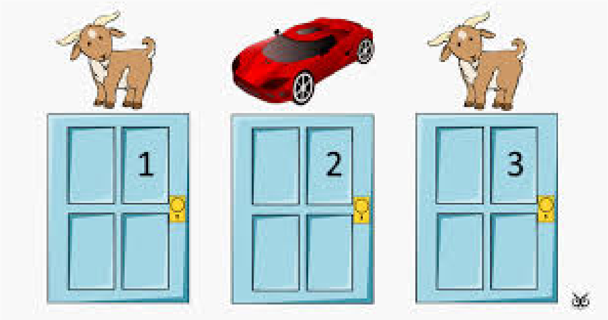
Annual income	Took 209	Took 309	TOTAL
Under \$50,000	36	24	60
\$50,000 to \$100,000	109	56	165
over \$100,000	35	40	75
Total	180	120	300

What is the probability of a having taken 309, conditional on making over \$100,000?

The Monty Hall Paradox!

What is the Monty Hall Paradox?

The Monty Hall Paradox!



What is the probability of winning a car when not switching?

$P(\text{car}) = ?$

The Monty Hall Paradox!

What is the probability of winning a car when not switching?

$P(\text{car}) = \frac{1}{3}$

The Monty Hall Paradox!

What is the probability of winning a car when switching?

The Monty Hall Paradox!

What is the probability of winning a car when switching?

Consider two scenarios: picking door with car first and picking door with goat first

The Monty Hall Paradox: switching

Consider two scenarios: picking door with car first and picking door with goat first

1. What is the probability of getting the car when switching after picking the car first?
2. What is the probability of getting the car when switching after picking a goat first?

The Monty Hall Paradox: switching

P(car when switching) = P(car | car first)$\times$ P(car first) + P(car | goat first) $\times$ P(goat first)

The Monty Hall Paradox: switching

P(car when switching) = P(car | car first)$\times$ P(car first) + P(car | goat first) $\times$ P(goat first)

P(car when switching) = 0 $\times \frac{1}{3}$ + 1 $\times \frac{2}{3}$

The Monty Hall Paradox: switching

P(car when switching) = P(car | car first)$\times$ P(car first) + P(car | goat first) $\times$ P(goat first)

P(car when switching) = 0 $\times \frac{1}{3}$ + 1 $\times \frac{2}{3}$

P(car when switching) = $\frac{2}{3}$

The Monty Hall Paradox: in *R*

```
sims <- 1000
doors <- c("goat", "goat", "car")
result.switch <- result.noswitch <- rep(NA, sims)
for (i in 1:sims) {
    ## randomly choose the initial door
first <- sample(1:3, size = 1)
result.noswitch[i] <- doors[first]
remain <- doors[-first] # remaining two doors
## Monty chooses one door with a goat
monty <- sample((1:2)[remain == "goat"], size = 1)
result.switch[i] <- remain[-monty]
}
mean(result.noswitch == "car")
mean(result.switch == "car")
```

Bayes' Rule/Theorem

How should we update our beliefs about event A after learning about some data related to the event?

Bayes' Rule/Theorem

How should we update our beliefs about event A after learning about some data related to the event?

Example: What is the probability of a person developing lung cancer?

Bayes' Rule/Theorem

How should we update our beliefs about event A after learning about some data related to the event?

Example: What is the probability of a person developing lung cancer?

How does the probability change once we learn about the person's smoking habits?

Bayes Rule/Theorem

P(A | B) = $\frac{P(B|A)P(A)}{P(B)}$

P(A) : prior probability of event A

P(A | B): posterior probability of event A given observed data B

Bayes Rule/Theorem

$\mathrm{P}(\mathrm{A} \mid \mathrm{B}) = \frac{P(B|A)P(A)}{P(B)}$

P(A) : prior probability of event A

P(A | B): posterior probability of event A given observed data B

P(B | A): probability of observing B given A

Bayes Rule/Theorem

$\mathrm{P}(\mathrm{A} \mid \mathrm{B}) = \frac{P(B|A)P(A)}{P(B)}$

P(A) : prior probability of event A

P(A | B): posterior probability of event A given observed data B

P(B | A): probability of observing B given A

P(B | A) $\times$ P(A) ?

Bayes Rule/Theorem

$\mathrm{P(A \mid B)} = \frac{P(B|A)P(A)}{P(B)}$

P(A) : prior probability of event A

P(A | B): posterior probability of event A given observed data B

Bayes Rule/Theorem

$\mathrm{P(A \mid B)} = \frac{P(B|A)P(A)}{P(B)}$

P(A) : prior probability of event A

P(A | B): posterior probability of event A given observed data B

P(B | A): probability of observing B given A

Bayes Rule/Theorem

P(A | B) = $\frac{P(B|A)P(A)}{P(B)}$

P(A) : prior probability of event A

P(A | B): posterior probability of event A given observed data B

P(B | A): probability of observing B given A

P(B | A) × P(A) = P(B and A)

Bayes Rule/Theorem

$$\text{P(A} \mid \text{B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(BandA)}{P(B)}$$

Bayes Rule/Theorem

$$\text{P}(\text{A} \mid \text{B}) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$$

Bayes Rule/Theorem

- Does your doctor know Bayes' rule? Cause he/she should!
- Example of medical tests:
  - every test comes with a reliability/accuracy
  - remember: false positive, false negative, etc

Bayes Rule/Theorem

What is the probability of being pregnant, given that you have a positive test?

P(p | +) = $\frac{P(+|p)P(p)}{P(+)}$

Bayes Rule/Theorem

What is the probability of being pregnant, given that you have a positive test?

P(p | +) = $\frac{P(+|p)P(p)}{P(+)}$

Decompose P(+), say test is 99 % accurate

Bayes Rule/Theorem

What is the probability of being pregnant, given that you have a positive test?

$$\text{P(preg} \mid + \text{)} = \frac{P(+|p)P(p)}{P(+)} = \frac{P(+|p)P(p)}{P(+|p)P(p)+P(+|\text{not p})P(\text{not p})}$$

Bayes Rule/Theorem

What is the probability of being pregnant, given that you have a positive test?

$$P(p \mid +) = \frac{P(+|p)P(p)}{P(+|p)P(p)+P(+|\text{not p})P(\text{not p})}$$

Bayes Rule/Theorem

What is the probability of being pregnant, given that you have a positive test?

$$P(p \mid +) = \frac{P(+|p)P(p)}{P(+|p)P(p)+P(+|\text{not p})P(\text{not p})}$$

$$P(p \mid +) = \frac{0.99P(p)}{0.99P(p)+0.05P(\text{not p})}$$

Bayes Rule/Theorem

What is the probability of being pregnant, given that you have a positive test?

$$\mathrm{P(p \mid +)} = \frac{P(+|p)P(p)}{P(+|p)P(p)+P(+|\text{not p})P(\text{not p})}$$

$$\mathrm{P(p \mid +)} = \frac{0.99P(p)}{0.99P(p)+0.05P(\text{not p})}$$

$$\mathrm{P(p \mid +)} = \frac{0.990.5}{0.99\times0.5+0.05\times0.5} = 0.95$$

Bayes Rule/Theorem

But what happens if your prior probability is stronger?

Bayes Rule/Theorem

But what happens if your prior probability is stronger?

$\text{P}(\text{p} \mid +) = \frac{0.99P(p)}{0.99P(p)+0.05P(\text{not p})}$

Bayes Rule/Theorem

But what happens if your prior probability is stronger?

$\mathrm{P(p \mid +)} = \frac{0.99P(p)}{0.99P(p)+0.05P(\text{not p})}$

$\mathrm{P(p \mid +)} = \frac{0.990.2}{0.99\times0.2+0.05\times0.8} = 0.83$

Bayes Rule/Theorem

But what happens if your prior probability is stronger?

$\mathrm{P}(\mathrm{p} \mid +) = \frac{0.990.05}{0.99 \times 0.05 + 0.01 \times 0.95} = 0.51$

Bayes Rule/Theorem

Many of our tests are not this good and disease is very rare:

- high-risk for down syndrome test
- $P(+ \mid DS) = 0.86$
- $P(+ \mid \text{not } DS) = 0.05$
- $P(DS) = 0.003$

Bayes Rule/Theorem

Many of our tests are not this good and disease is very rare:

- high-risk for down syndrome test
- P(+ | DS) = 0.86
- P(+ | not DS) = 0.05
- P(DS) = 0.003

$$\mathrm{P}(\mathrm{DS} \mid +) = \frac{P(+|DS)P(DS)}{P(+|DS)P(DS)+P(+|\text{not DS})P(\text{not DS})}$$

Bayes Rule/Theorem

Many of our tests are not this good and disease is very rare:

- high-risk for down syndrome test
- P(+ | DS) = 0.86
- P(+ | not DS) = 0.05
- P(DS) = 0.003

$$\text{P(DS} \mid +) = \frac{P(+|DS)P(DS)}{P(+|DS)P(DS)+P(+|\text{not DS})P(\text{not DS})}$$

$$\text{P(DS} \mid +) = \frac{0.86 \times 0.003}{0.86 \times 0.003 + 0.05 * 0.997}$$

Bayes Rule/Theorem

Many of our tests are not this good and disease is very rare:

- high-risk for down syndrome test
- $P(+ \mid DS) = 0.86$
- $P(+ \mid \text{not } DS) = 0.05$
- $P(DS) = 0.003$

Bayes Rule/Theorem

Many of our tests are not this good and disease is very rare:

- high-risk for down syndrome test
- P(+ | DS) = 0.86
- P(+ | not DS) = 0.05
- P(DS) = 0.003

$$\text{P(DS} \mid +) = \frac{P(+|DS)P(DS)}{P(+|DS)P(DS)+P(+|\text{not DS})P(\text{not DS})}$$

Bayes Rule/Theorem

Many of our tests are not this good and disease is very rare:

- high-risk for down syndrome test
- P(+ | DS) = 0.86
- P(+ | not DS) = 0.05
- P(DS) = 0.003

$$\mathrm{P(DS} \mid +) = \frac{P(+|DS)P(DS)}{P(+|DS)P(DS)+P(+|\text{not DS})P(\text{not DS})}$$

$$\mathrm{P(DS} \mid +) = \frac{0.86 \times 0.003}{0.86 \times 0.003 + 0.05 * 0.997} = 0.049$$

Changes with age!

- What is a random variable? We assigns a number to an event
    - coin flip: tail$=0$; heads$=1$
    - Senate election: Ted Cruz$=0$; Beto O'Rourke$=1$
    - Voting: vote $=1$; not vote $=0$

Random Variables and Probability Distributions

- What is a random variable? We assigns a number to an event
  - coin flip: tail= 0; heads= 1
  - Senate election: Ted Cruz= 0; Beto O'Rourke= 1
  - Voting: vote = 1; not vote = 0

Probability distribution: Probability of an event that a random variable takes a certain value

- P(coin =1); P(coin = 0)
- P(election = 1); P(election = 0)

Random Variables and Probability Distributions

- Probability density function (PDF): f(x) How likely does X take a particular value?
- Probability mass function (PMF): When X is discrete, f(x)=P(X =x)

Random Variables and Probability Distributions

- Probability density function (PDF): f(x) How likely does X take a particular value?
- Probability mass function (PMF): When X is discrete, $f(x)=P(X=x)$
- Cumulative distribution function (CDF): $F(x) = P(X \leq x)$
    - What is the probability that a random variable X takes a value equal to or less than x?
    - Area under the density curve (either we use the sum Σ or integral $\int$)
    - Non-decreasing

Random Variables and Probability Distributions: Binomial Distribution

- PMF: for $x \in \{0, 1, \ldots, n\}$,
  $f(x) \ = \ P(X = x) \ = \ \binom{n}{x} p^x (1-p)^{n-x}$
- CDF: for $x \in \{0, 1, \ldots, n\}$
  $F(x) \ = \ P(X \leq x) \ = \ \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$

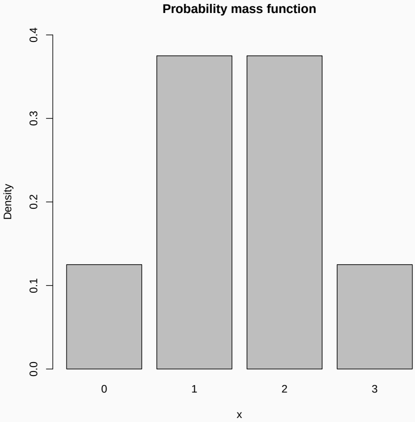
Random Variables and Probability Distributions: Binomial Distribution

- Example: flip a fair coin 3 times
  $$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$
  $$f(x) = P(X = 1) = \binom{3}{1} 0.5^1 (0.5)^2 = 3 * 0.5 * 0.5^2 = 0.375$$

Random Variables and Probability Distributions: Binomial Distribution

```
x <- 0:3
barplot(dbinom(x, size = 3, prob = 0.5), ylim = c(0, 0.4), names.arg = x, xlab = "x",
        ylab = "Density", main = "Probability mass function")
```

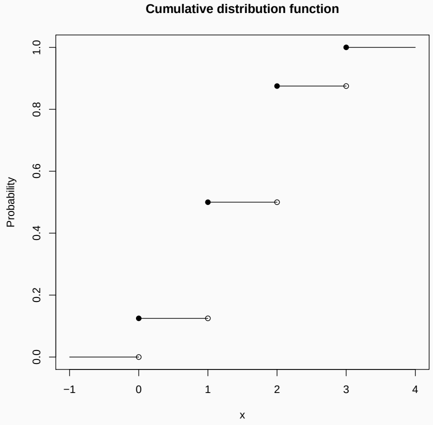
Random Variables and Probability Distributions: Binomial Distribution



Random Variables and Probability Distributions: Binomial Distribution

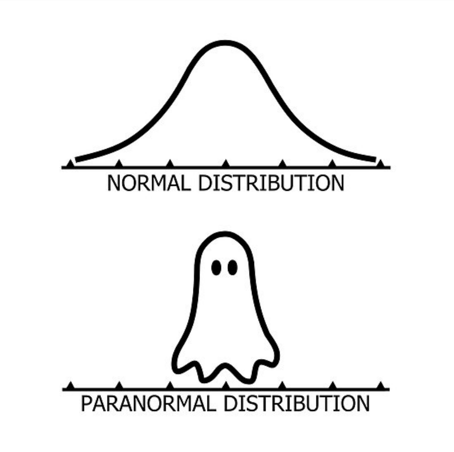
```
x <- -1:4
pb <- pbinom(x, size = 3, prob = 0.5)
plot(x[1:2], rep(pb[1], 2), ylim = c(0, 1), type = "s", xlim = c(-1, 4), xlab = "x",
     ylab = "Probability", main = "Cumulative distribution function")
for (i in 2:(length(x)-1)) {
    lines(x[i:(i+1)], rep(pb[i], 2))
}
points(x[2:(length(x)-1)], pb[2:(length(x)-1)], pch = 19)
points(x[2:(length(x)-1)], pb[1:(length(x)-2)])
```

Random Variables and Probability Distributions: Binomial Distribution



Random Variables and Probability Distributions: Normal Distribution

Normal distribution



Random Variables and Probability Distributions: Normal Distribution

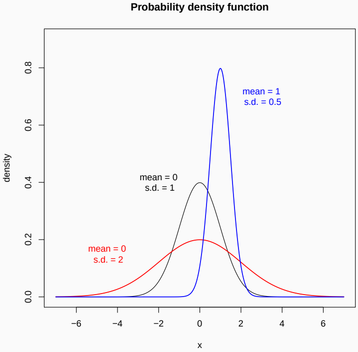
Normal distribution also called Gaussian distribution

Random Variables and Probability Distributions: Normal Distribution

Normal distribution with mean μ and standard deviation σ

- PDF: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
- CDF (no simple formula. use to compute it):
  $F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dt$

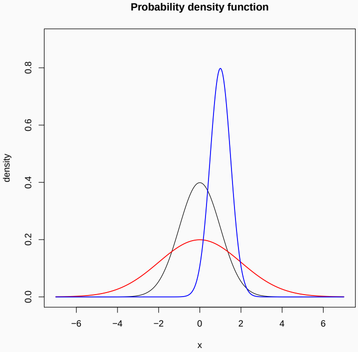
Random Variables and Probability Distributions: Normal Distribution



Random Variables and Probability Distributions: Normal Distribution in R

```
x <- seq(from = -7, to = 7, by = 0.01)
plot(x, dnorm(x), xlab = "x", ylab = "density", type = "l",
     main = "Probability density function", ylim = c(0, 0.9))
lines(x, dnorm(x, sd = 2), col = "red", lwd = lwd)
lines(x, dnorm(x, mean = 1, sd = 0.5), col = "blue", lwd = lwd)
```

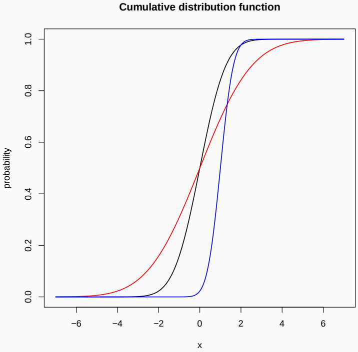
Random Variables and Probability Distributions: Normal Distribution in R



Random Variables and Probability Distributions: Normal Distribution in R

```
plot(x, pnorm(x), xlab = "x", ylab = "probability", type = "l",
     main = "Cumulative distribution function", lwd = lwd)
lines(x, pnorm(x, sd = 2), col = "red", lwd = lwd)
lines(x, pnorm(x, mean = 1, sd = 0.5), col = "blue", lwd = lwd)
```

Random Variables and Probability Distributions: Normal Distribution in R



Random Variables and Probability Distributions: Normal Distribution

Let $X \sim N(\mu, \sigma^2)$, and c be some constant

- Adding/subtracting to/from a random variable that is normally distributed also results in a variable with a normal distribution: $\mathrm{Z} = \mathrm{X} + \mathrm{c}$ then $Z \sim N(\mu + c, \sigma^2)$

Let $X \sim N(\mu, \sigma^2)$, and c be some constant

- Adding/subtracting to/from a random variable that is normally distributed also results in a variable with a normal distribution: Z = X + c then $Z \sim N(\mu + c, \sigma^2)$
- Multiplying or dividing a random variable that is normally distributed also results in a variable with a normal distribution: $Z = X \times c$ then $Z \sim N(\mu \times c, (\sigma \times c)^2)$
- Z-score of a random variable that is normally distributed has mean 0 and sd = 1

Random Variables and Probability Distributions: Normal Distribution

Curve of the standard normal distribution:

- Symmetric around 0
- Total area under the curve is 100%
- Area between -1 and 1 is ~68%
- Area between -2 and 2 is ~95%
- Area between -3 and 3 is ~99.7%

Random Variables and Probability Distributions: Normal Distribution

```
x <- seq(from = -7, to = 7, by = 0.01)
lwd <- 1.5
plot(x, dnorm(x), xlab = "x", ylab = "density", type = "l",
     main = "Probability density function", ylim = c(0, 0.9))
abline(v= -1, col = "red")
abline(v= 1, col = "red")
abline(v= -2, col = "green")
abline(v= 2, col = "green")
```

Random Variables and Probability Distributions: Normal Distribution