Week 10: Sampling Distributions and Limits

MATH697

Sahir Bhatnagar

November 7, 2017

McGill University

Sampling Distributions and Limits

Preview

- This section make the **transition between probability and inferential statistics**.

Preview

- This section make the **transition between probability and inferential statistics**.
- Given a sample of n observations from a population, we will be calculating estimates of the population mean, median, standard deviation, and various other population characteristics (parameters).

Preview

- This section make the **transition between probability and inferential statistics**.
- Given a sample of n observations from a population, we will be calculating estimates of the population mean, median, standard deviation, and various other population characteristics (parameters).
- Prior to obtaining data, there is uncertainty as to which of all possible samples will occur.

Preview

- This section make the **transition between probability and inferential statistics**.
- Given a sample of n observations from a population, we will be calculating estimates of the population mean, median, standard deviation, and various other population characteristics (parameters).
- Prior to obtaining data, there is uncertainty as to which of all possible samples will occur.
- Because of this, estimates such as $\bar{x}$ (the sample mean) will vary from one sample to another

Preview

- The behavior of such estimates in repeated sampling is described by what are called **sampling distributions**.

Preview

- The behavior of such estimates in repeated sampling is described by what are called **sampling distributions**.
- Any particular sampling distribution will give an indication of how close the estimate is likely to be to the value of the parameter being estimated.

- We will use probability results to study sampling distributions.

Preview

- We will use probability results to study sampling distributions.
- A particularly important result is the **Central Limit Theorem**, which shows how the behavior of the sample mean can be described by a particular normal distribution when the sample size is large.

Statistics and Their Distributions

Two random samples will be different

- The observations in a single sample are denoted by $x_1, x_2, \ldots, x_n$

Two random samples will be different

- The observations in a single sample are denoted by $x_1, x_2, \ldots, x_n$
- Consider selecting two different samples of size n from the same population distribution.

Two random samples will be different

- The observations in a single sample are denoted by $x_1, x_2, \ldots, x_n$
- Consider selecting two different samples of size n from the same population distribution.
- The x_i's in the second sample will virtually always differ at least a bit from those in the first sample.

Uncertainty in Summary Measures of the Random Samples

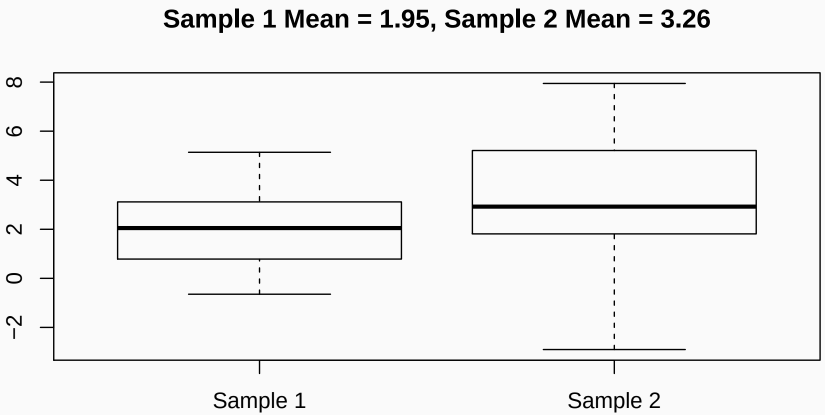
- This variation in observed values in turn implies that the value of any function of the sample observations - such as the sample mean or sample standard deviation also varies from sample to sample.

Uncertainty in Summary Measures of the Random Samples

- This variation in observed values in turn implies that the value of any function of the sample observations - such as the sample mean or sample standard deviation also varies from sample to sample.
- That is, prior to obtaining $x_1, \ldots, x_n$, there uncertainty as to the value of $\overline{x}$ and s (the sample standard deviation)

Two Random Samples from a N(2,4) Distribution

```
x1 <- rnorm(10, 2, 2) ; x2 <- rnorm(10, 2, 2)
boxplot(x1,x2, main = sprintf("Sample 1 Mean = %0.2f, Sample 2 Mean = %0.2f",
        mean(x1), mean(x2)), names = c("Sample 1","Sample 2"))
```



A Statistic

Definition 1 (Statistic)

- A **statistic** is any quantity whose value can be calculated from sample data
- Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result.
- A **statistic is a random variable** and will be denoted by an uppercase letter (e.g. $\bar{X}$)
- A lowercase letter is used to represent the calculated or observed value of the statistic (e.g. $\bar{x}$)

- Suppose a drug is given to a sample of patients, another drug is given to a second sample, and the cholesterol levels are denoted by $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively.

Sample Mean is a Statistic

- Suppose a drug is given to a sample of patients, another drug is given to a second sample, and the cholesterol levels are denoted by $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively.
- The statistic $\overline{X} - \overline{Y}$, i.e., the difference between the two sample mean cholesterol levels, may be important.

Any Statistic has a Probability Distribution

- Suppose, for example, that $n = 2$ components are randomly selected and the number of breakdowns while under warranty is determined for each one.

Any Statistic has a Probability Distribution

- Suppose, for example, that $n = 2$ components are randomly selected and the number of breakdowns while under warranty is determined for each one.
- Possible values for the sample mean number of breakdowns $\overline{X}$ are

Any Statistic has a Probability Distribution

- Suppose, for example, that $n = 2$ components are randomly selected and the number of breakdowns while under warranty is determined for each one.
- Possible values for the sample mean number of breakdowns $\overline{X}$ are

Any Statistic has a Probability Distribution

- Suppose, for example, that $n = 2$ components are randomly selected and the number of breakdowns while under warranty is determined for each one.
- Possible values for the sample mean number of breakdowns $\bar{X}$ are

X_1	X_2	$\bar{X}$
0	0	0
0	1	0.5
1	0	0.5
0	2	1
2	0	1
$\vdots$	$\vdots$	$\vdots$

Probability Distribution of Statistic is its Sampling Distribution

- The probability distribution of $\bar{X}$ specifies $P\big(\bar{X} = 0\big)$, $P\big(\bar{X} = 0.5\big)$, $P\big(\bar{X} = 1\big)$ and so on

Probability Distribution of Statistic is its Sampling Distribution

- The probability distribution of $\bar{X}$ specifies $P(\bar{X} = 0)$, $P(\bar{X} = 0.5)$, $P(\bar{X} = 1)$ and so on
- From these, other probabilities such as $P(1 \leq \bar{X} \leq 3)$ and $P(\bar{X} \geq 2.5)$ can be calculated

Probability Distribution of Statistic is its Sampling Distribution

- The probability distribution of $\bar{X}$ specifies $P(\bar{X} = 0)$, $P(\bar{X} = 0.5)$, $P(\bar{X} = 1)$ and so on
- From these, other probabilities such as $P(1 \leq \bar{X} \leq 3)$ and $P(\bar{X} \geq 2.5)$ can be calculated
- The probability distribution of a statistic is referred to as its **sampling distribution** to emphasize that it describes how the statistic varies in value across all samples that might be selected.

Definition 2 (Random Sample)

The random variables $X_1, X_2, \ldots, X_n$ are said to form a **random sample** of size n is

- The X_i's are independent random variables
- Every X_i has the same probability distribution

These two conditions can be paraphrased by saying that the X_i's are *independent and identically distributed* (iid).

Deriving the Sampling Distribution of a Statistic

- Probability rules can be used to obtain the distribution of a statistic provided that it is a fairly simple function of the X_i's and either there are relatively few different X values in the population or else the population distribution has a nice form

Deriving the Sampling Distribution of a Statistic

- Probability rules can be used to obtain the distribution of a statistic provided that it is a fairly simple function of the X_i's and either there are relatively few different X values in the population or else the population distribution has a nice form
- The next examples illustrate such a situation and provides a motivation for finding an approximation of the sampling distribution

Example 3 (MP3 Players)

A certain brand of MP3 player comes in three configurations:

memory	2 GB	4 GB	8 GB
x (cost)	80	100	120
$p(x)$	0.20	0.30	0.50

With $\mu = 106, \sigma^2 = 244$. Suppose only two MP3 players are sold today: X_1 and X_2 representing the cost of the 1st and 2nd player, respectively. When $n = 2$, $s^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2$

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
80	80	$(.2)(.2) = .04$	80	0
80	100	$(.2)(.3) = .06$	90	200
80	120	$(.2)(.5) = .10$	100	800
100	80	$(.3)(.2) = .06$	90	200
100	100	$(.3)(.3) = .09$	100	0
100	120	$(.3)(.5) = .15$	110	200
120	80	$(.5)(.2) = .10$	100	800
120	100	$(.5)(.3) = .15$	110	200
120	120	$(.5)(.5) = .25$	120	0

Example 4 (MP3 Players)

To obtain the probability distribution of $\overline{X}$, the sample average cost per MP3 player, we must consider each possible value $\bar{x}$ and compute its probability, e.g., $P(\bar{x} = 100) = 0.10 + 0.09 + 0.10 = 0.29$, $P(S^2 = 800) = 0.10 + 0.10 = 0.20$. The complete sampling distributions of $\overline{X}$ and S^2 are given below:

$\bar{x}$	80	90	100	110	120
$p_{\overline{X}}(\bar{x})$	.2	.12	.29	.30	.5

s^2	0	200	800
$p_{S^2}(s^2)$	.38	.42	.20

- $E(\overline{X}) = \sum \bar{x} p_{\overline{X}}(\bar{x}) = 106 = \mu$
- $V(\overline{X}) = \sum (\bar{x} - \mu)^2 = \sum (\bar{x} - 106)^2 p_{\overline{X}}(\bar{x}) = 122 = 244/2 = \sigma^2/2$ (half the population variance: why?)
- $E(S^2) = \sum s^2 p_{S^2}(s^2) = 0(0.38) + 200(0.42) + 800(0.20) = 244 = \sigma^2$

Example 5 (MP3 Players)

The probability histogram for both the original distribution X (a) and the $\overline{X}$ (b) distribution. We see that the mean of $\overline{X}$ (denoted by $E(\overline{X})$) is equal to the mean of the original distribution. We also see that the $\overline{X}$ distribution has **smaller spread** than the original distribution, since the values of $\overline{x}$ are **more concentrated toward the mean**. The $\overline{X}$ sampling distribution is centered at the population mean μ.

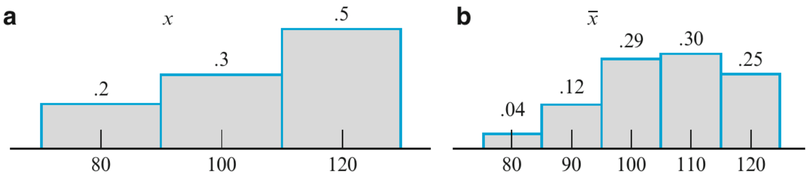


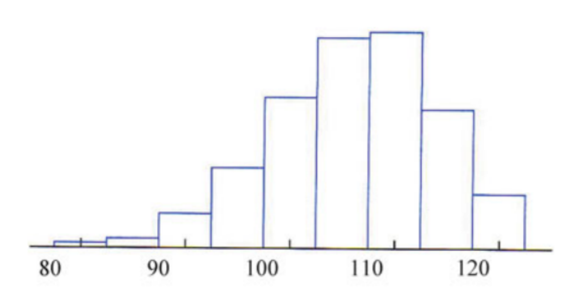
Figure 6.2 Probability histograms for (a) the underlying population distribution and (b) the sampling distribution of $\overline{X}$ in Example 6.2

Example 6 (MP3 Players)

If four MP3 players had been purchased on the day of interest, the sample average cost $\overline{X}$ would be based on a random sample of four X_i's. More calculation eventually yields the distribution of $\overline{X}$ for $n = 4$ as

$\bar{x}$	80	85	90	95	100	105	110	115	120
$p_{\overline{X}}(\bar{x})$	.0016	.0096	.0376	.0936	.1761	.2340	.2350	.1500	.0625

From this, $E(\overline{X}) = 106 = \mu$ and $V(\overline{X}) = 61 = \sigma^2/4$



Some Remarks

- The previous example showed us that the computation of $p_{\overline{X}}(\overline{x})$ and $p_{S^2}(s^2)$ can be tedious

Some Remarks

- The previous example showed us that the computation of $p_{\bar{X}}(\bar{x})$ and $p_{S^2}(s^2)$ can be tedious
- This example should also suggest that there are some general relationships between $E(\bar{X})$, $V(\bar{X})$, $E(S^2)$ and the population mean μ and variance σ^2.

Some Remarks

- The previous example showed us that the computation of $p_{\bar{X}}(\bar{x})$ and $p_{S^2}(s^2)$ can be tedious
- This example should also suggest that there are some general relationships between $E(\bar{X})$, $V(\bar{X})$, $E(S^2)$ and the population mean μ and variance σ^2.
- Sampling distributions can sometimes be computed by direct computation or by approximations such as the central limit theorem (CLT)

Some Remarks

- The previous example showed us that the computation of $p_{\bar{X}}(\bar{x})$ and $p_{S^2}(s^2)$ can be tedious
- This example should also suggest that there are some general relationships between $E(\bar{X})$, $V(\bar{X})$, $E(S^2)$ and the population mean μ and variance σ^2.
- Sampling distributions can sometimes be computed by direct computation or by approximations such as the central limit theorem (CLT)
- Techniques for deriving such approximations will be discussed next

Convergence in Probability

Definition 7 (Convergence in Probabilty)

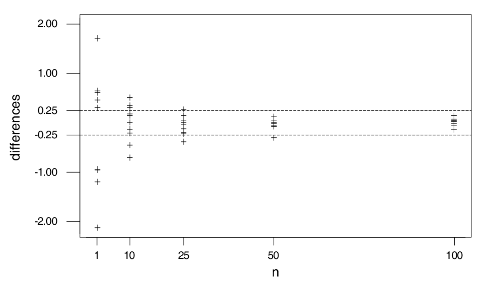
Let $X_1, X_2, \ldots$ be an infinite sequence of random variables, and let Y be another random variable. Then the sequence $\{X_n\}$ converges in probability to Y, if for all $\epsilon > 0$

$$\lim_{n \to \infty} P(|X_n - Y| \geq \epsilon) = 0 \tag{1}$$

Alternatively we write $X_n \xrightarrow{p} Y$

Convergence in Probability

We plot the differences $X_n - Y$ for selected values of n, for 10 generated sequences $\{X_n - Y\}$ for a typical situation where the random variables X_n converge to a random variable Y in probability. We have also plotted the horizontal lines at $\pm\epsilon$ for $\epsilon = 0.25$. From this we can see the increasing concentration of the distribution of $X_n - Y$ about 0, as n increases, as required by Definition (7). In fact, the 10 observed values of $ X_{100} - Y$ all satisfy the inequality $\left|X_{100} - Y\right| < 0.25$.



Convergence in Probability Example

Example 8 (Identical Random Variables)

Let Y be any random variable, and let $X_1 = X_2 = X_3 = \cdots = Y$, i.e., the random variables are all identical to each other.

Convergence in Probability Example

Example 9 (Functions of Uniforms)

Let $U \sim Uniform(0,1)$. Define X_n by

$$X_n = \begin{cases} 3 & U \leq 2/3 - 1/n \\ 8 & otherwise \end{cases}$$

and define Y by

$$Y = \begin{cases} 3 & U \leq 2/3 \\ 8 & otherwise \end{cases}$$

Convergence in Probability Example

Example 10 (Exponential and a Constant)

Let $Z_n \sim \mathit{Exponential}(n)$ and let $Y = 0$.

Important application of convergence in probability

- One of the most important applications of convergence in probability is the **weak law of large numbers**

Important application of convergence in probability

- One of the most important applications of convergence in probability is the **weak law of large numbers**
- Suppose $X_1, X_2, \cdots$ is a sequence of independent random variables that each have the same mean μ and variance σ^2

Important application of convergence in probability

- One of the most important applications of convergence in probability is the **weak law of large numbers**
- Suppose $X_1, X_2, \cdots$ is a sequence of independent random variables that each have the same mean μ and variance σ^2
- For all positive integers n, let

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

be the *sample average*, or *sample mean*, for $X_1, \cdots, X_n$

Important application of convergence in probability

- One of the most important applications of convergence in probability is the **weak law of large numbers**
- Suppose $X_1, X_2, \cdots$ is a sequence of independent random variables that each have the same mean μ and variance σ^2
- For all positive integers n, let

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

  be the *sample average*, or *sample mean*, for $X_1, \cdots, X_n$
- When the sample size n is fixed, we will often use $\bar{X}$ as a notation for sample mean instead of $\bar{X}_n$.

Important application of convergence in probability

- One of the most important applications of convergence in probability is the **weak law of large numbers**
- Suppose $X_1, X_2, \cdots$ is a sequence of independent random variables that each have the same mean μ and variance σ^2
- For all positive integers n, let

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

  be the *sample average*, or *sample mean*, for $X_1, \cdots, X_n$
- When the sample size n is fixed, we will often use $\bar{X}$ as a notation for sample mean instead of $\bar{X}_n$.
- The sample mean is itself a random variable with mean μ and variance σ^2/n (why?)

Coin Flips

- If we flip a sequence of fair coins, and if $X_i = 1$ or $X_i = 0$ as the ith coin comes up heads or tails, then $\bar{X}_n$ represents the fraction of the first n coins that came up heads

Coin Flips

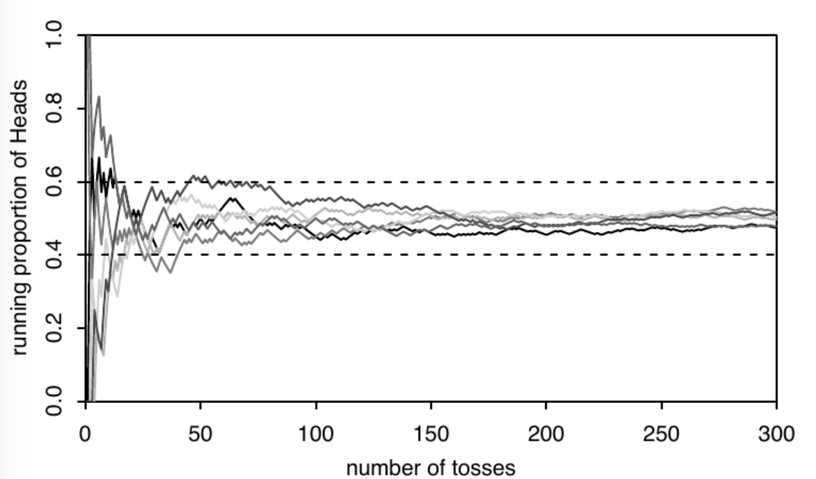
- If we flip a sequence of fair coins, and if $X_i = 1$ or $X_i = 0$ as the ith coin comes up heads or tails, then $\overline{X}_n$ represents the fraction of the first n coins that came up heads
- We might expect that for large n, this fraction will be close to $1/2$, i.e., to the expected value of the X_i

Coin Flips

- If we flip a sequence of fair coins, and if $X_i = 1$ or $X_i = 0$ as the ith coin comes up heads or tails, then $\overline{X}_n$ represents the fraction of the first n coins that came up heads
- We might expect that for large n, this fraction will be close to $1/2$, i.e., to the expected value of the X_i
- The **weak law of large numbers** provides a precise sense in which average values $\overline{X}_n$ tend to get close to $E(X_i)$, for large n

Coin Flips

- Running proportion of Heads in 6 sequences of fair coin tosses. Dashed lines at 0.6 and 0.4 are plotted for reference. As the number of tosses increases, the proportion of Heads approaches 1/2.



Weak Law of Large Numbers

Theorem 11 (Weak Law of Large Numbers (WLLN))

Let $X_1, X_2, \cdots$, be a sequence of independent random variables, each having the same mean μ and each having variance less than or equal to $\nu < \infty$. Then for all $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(|\bar{X}_n - \mu| \geq \epsilon\right) = 0 \tag{2}$$

That is, the averages converge in probability to the common mean μ or $\bar{X}_n \xrightarrow{p} \mu$

Proof: on board

Example 12 (Fair coins)

Consider flipping a sequence of identical fair coins. Let $\bar{X}_n$ be the fraction of the first n coins that are heads. Then $\bar{X}_n = (X_1 + \cdots + X_n)/n$, where $X_i = 1$ if the ith coin is heads, otherwise $X_i = 0$.

Example 13 (Normal RVs)

Let $X_1, X_2, \ldots$ be iid with distribution $N(3, 5)$.

WLLN Importance

- The law of large numbers (LLN) is essential for **simulations, statistics, and science!!**.

WLLN Importance

- The law of large numbers (LLN) is essential for **simulations, statistics, and science!!**.
- Consider generating data from a large number of independent replications of an experiment, performed either by computer simulation or in the real world

WLLN Importance

- The law of large numbers (LLN) is essential for **simulations, statistics, and science!!**.
- Consider generating data from a large number of independent replications of an experiment, performed either by computer simulation or in the real world
- Every time we use the proportion of times that something happened as an approximation to its probability, we are implicitly appealing to LLN.

WLLN Importance

- The law of large numbers (LLN) is essential for **simulations, statistics, and science!!**.
- Consider generating data from a large number of independent replications of an experiment, performed either by computer simulation or in the real world
- Every time we use the proportion of times that something happened as an approximation to its probability, we are implicitly appealing to LLN.
- Every time we use the average value in the replications of some quantity to approximate its theoretical average, we are implicitly appealing to LLN.

Summary

- A sequence $\{X_n\}$ of random variables converges in probability to Y if

$$\lim_{n\to\infty} P\big(|X_n - Y| \geq \epsilon\big) = 0$$

Summary

- A sequence $\{X_n\}$ of random variables converges in probability to Y if

$$\lim_{n\to\infty} P\big(|X_n - Y| \geq \epsilon\big) = 0$$

- The Weak Law of Large Numbers (WLLN) says that if $\{X_n\}$ is iid, then

$$\bar{X}_n = \big(X_1 + \cdots + X_n\big)/n \xrightarrow{p} E(X_i)$$

Convergence with Probability 1

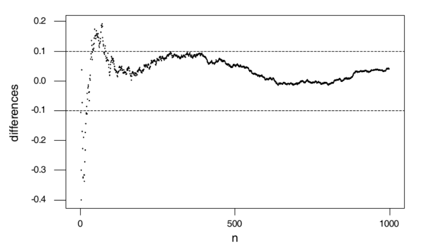
Definition 14 (Convergence with Probability 1)

Let $X_1, X_2, \ldots,$ be an infinite sequence of random variables. We shall say that the sequence $\{X_i\}$ **converges with probability 1** (or **converges almost surely** (a.s)) to a random variable Y if

$$P(\lim_{n\to\infty} X_n = y) = 1 \tag{3}$$

we write this as $X_n \xrightarrow{a.s.} Y$

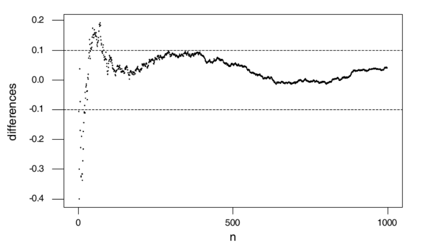
- Graph of the sequence of differences $\{X_n - Y\}$ for a typical situation where the random variables X_n converge to a random variable Y with probability 1.



- Definition (14) indicates that for any given $\varepsilon > 0$, there will exist a value N_ε such that $|X_n - Y| < \varepsilon$ for every $n \geq N_\varepsilon$.

Convergence with Probability 1

- Graph of the sequence of differences $\{X_n - Y\}$ for a typical situation where the random variables X_n converge to a random variable Y with probability 1.



- Definition (14) indicates that for any given $\varepsilon > 0$, there will exist a value N_ε such that $|X_n - Y| < \varepsilon$ for every $n \geq N_\varepsilon$.
- Contrast this with the situation depicted for convergence in probability, which only says that the probability distribution $X_n - Y$ concentrates about 0 as n grows and not that the individual values of $X_n - Y$ will necessarily all be near 0

Session Info

```
devtools::session_info()
```

```
##  setting  value
##  version  R version 3.4.1 (2017-06-30)
##  system   x86_64, linux-gnu
##  ui       X11
##  language en_US
##  collate  en_US.UTF-8
##  tz       Canada/Eastern
##  date     2017-11-07
##
##  package     * version    date       source
##  abind         1.4-5      2016-07-21 cran (@1.4-5)
##  arm           1.9-3      2016-11-27 cran (@1.9-3)
##  assertthat    0.2.0      2017-04-11 CRAN (R 3.4.1)
##  backports     1.1.0      2017-05-22 cran (@1.1.0)
##  base        * 3.4.1      2017-07-08 local
##  bindr         0.1        2016-11-13 CRAN (R 3.4.1)
##  bindrcpp      0.2        2017-06-17 CRAN (R 3.4.1)
##  blme          1.0-4      2015-06-14 cran (@1.0-4)
##  broom         0.4.2      2017-02-13 CRAN (R 3.4.1)
```