Math 321 – Chapter 4 – The Central Limit Theorem
(draft version 2019/04/17-09:21:00)

Contents

1 Independence for random variables

Independence. Let X and Y be real-valued random variables. We say X and Y are independent is the events $\{a < X < b\}$ and $\{c < Y < d\}$ are independent for all a, b, c, d. Intuitively, this means that knowing the value of one of the variables doesn't affect the probabilities for the other variable.

Example: Let X be the outcome for a roll of a 6-sided die, and Y be the outcome for a roll of a 20-sided die. It should be clear that they are independent random variables. The probability that the first die is 5 and the second die is 18 can be calculated as

$$P(X = 5 \text{ and } Y = 18) = P(X = 5) \cdot P(Y = 18) = \frac{1}{6} \cdot \frac{1}{20}.$$

Identically distributed. Random variables that have the same probability distribution function are called *identically distributed*.

Example: Consider random variables X ={the numerical outcome of a fair, green 6-sided die} and Y =the numerical outcome of a fair, red 6-sided die}. These two random variables have the same possible values $\{1, 2, 3, 4, 5, 6\}$ and the same probabilities for each value; they are identically distributed.

Example: Consider a coin that is flipped a number of times. The outcome for each coin flip is identically distributed.

The two above examples are also independent. Random variable that are independent and identically distributed are called... *independent and identically distributed* or for short, iid.

2 Law of large numbers

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed (iid) random variables with mean μ. That is, $E(X_i) = \mu$ for $i = 1, 2, \ldots, n$. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \text{ converges in probability to } \mu \text{ as } n \to \infty.$$

Another way to phrase this is that for all positive numbers $\epsilon > 0$,

$$P\left(\left|\mu - \frac{1}{n}\sum_{i=1}^{n} X_i\right| > \epsilon\right) \longrightarrow 0 \text{ as } n \to \infty.$$

How to interpret this: Remember that the actual value of $\frac{1}{n}\sum_{i=1}^{n} X_i$ is still random, so it could vary quite a bit if we were to sample different X_i's. But as the sample size n gets very large, the chance of the sample average deviating too far from the mean μ is very low.

We have already seen this in class in several cases.

Example: Let's consider n coin flips with $P(H) = p$. Let X_i be the outcome of the i^{th} coin flip, i.e. $X_i = 1$ if the i^{th} coin flip is heads and $X_i = 0$ if it is tails. Then the X_i's are independent, and they are all Bernoulli random variables with probability of success p. So they are iid (independent and identically distributed). Thus if we take the average value of n coin flips: $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, as we increase n, this average will get very close to p (which is the μ in this case). We say that $\overline{X}_n$ converges to p in probability as $n \to \infty$. Recall that you observed this in one of the R activity assignments.

3 Central limit theorem

Let $X_1, X_2, \ldots, X_n$ be iid with mean μ and variance σ^2. Note that we are not assuming anything about their distribution other than the mean (expected value) and variance. They could be drawn from a normal distribution, exponential distribution, binomial, Poisson, or any other distribution that has a finite mean and variance.

Then when the sample size n is large, the sample mean will be approximately normally distributed with mean μ and variance σ^2/n.

Remember that $X_1, X_2, \ldots, X_n$ is a random sample. So for a fixed sample size n, we can draw a variety of different random samples. Each sample will have a different sample mean $\overline{X}_n$. So by drawing many many different samples, we will have many many different sample means. If we create a histogram for our dataset of sample means, it will look very much like the normal distribution with mean μ and variance σ^2/n.

This is called the central limit theorem (CLT).

This is a remarkable fact!!!! This is one of the reasons why the normal distribution is so important to all of statistics.

If the $X_1, X_2, \ldots, X_n$ are iid and $X_i \sim N(\mu, \sigma^2)$, in other words if they come from a normal distribution exactly, then $\overline{X}_n$ is exactly normally distributed.

Summary of CLT:

$$X_i \sim N(\mu, \sigma^2) \text{ for } i = 1, 2, \ldots, n \text{ then } \overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

If the X_i are not normally distributed, this still holds approximately when the sample size is sufficiently large.

How large should the sample size be so that the central limit theorem is a good approximation? In most cases $n \geq 30$ is a general rule of thumb. Even with smaller sample sizes, the approximation may not be too bad.

If the underlying distribution of the X_i is extremely skewed with a large probability for very far away outliers, then n in the 100s, 1000s or larger may be necessary. For example, if the X_i are exponentially distributed (a distribution that has a higher probability of large outliers), then a sample size of 100 or greater will be required for a decent approximation by the CLT.

3.1 Using the CLT to calculated probabilities for $\overline{X}_n$

Since we know that $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, we can now calculate the probability of the sample mean being in a particular range of values. It is now just a normal random variable!

Example: Human height is approximately normally distributed with mean 175 cm and standard deviation 7 cm. If 30 people are selected at random, approximate the probability that the mean height of the sample is greater than 182 cm.
Solution: $\overline{X}_{30} \sim N(\text{mean} = 175, \text{std. dev.} = 7/\sqrt{30})$.
Thus $P(\overline{X}_{30} \geq 182)$ =1-`pnorm(182,mean=175,sd=7/sqrt(30))`$\approx 2(10)^{-8}$. This is a minuscule probability! Note that the cut-off of 182 is only 1 standard deviation above the mean, so according to the 68-95-99.7 rule, there is a 17% chance of any individual data point being above 182, but the mean of a sample of 30 data points will rarely be that far away from 175.

4 Sampling Distributions

In discussing the central limit theorem, we are really talking about *sampling distributions*. A sampling distribution is the probability distribution of a sample statistic. A sample statistic is something that you calculate from a sample. For example the sample mean and sample variance are both two sample statistics. There are other sample statistics such as the sample median, sample skewness, and even many others.

Here is a graphical description of the process:

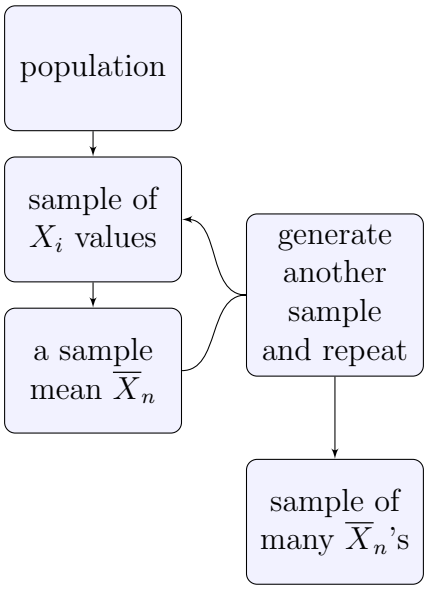

Figure 1: Gathering many sample means flow chart diagram.

If the X_i are normally distributed with mean μ and variance σ^2, then the sample mean is normally distributed with mean μ and variance $\frac{\sigma^2}{n}$ as stated by the CLT.

Here we will look at another example of a sampling distribution.

Consider a discrete finite population: $\{1, 1, 1, 1, 1, 2, 2, 2, 3, 3\}$. We will sample from this population with replacement. Let X_i be the i^{th} individual selected. We'll just select a sample of size $n = 2$.

Here are all possible samples of size 2, the probability of each, and the sample mean and variance for each sample:

	X_1	X_2	$P(X_1, X_2)$	$\overline{x}$	s^2
1	1.00	1.00	0.25	1.00	0.00
2	2.00	1.00	0.15	1.50	0.50
3	3.00	1.00	0.10	2.00	2.00
4	1.00	2.00	0.15	1.50	0.50
5	2.00	2.00	0.09	2.00	0.00
6	3.00	2.00	0.06	2.50	0.50
7	1.00	3.00	0.10	2.00	2.00
8	2.00	3.00	0.06	2.50	0.50
9	3.00	3.00	0.04	3.00	0.00

Then we look for each possible value of $\overline{x}$ and add up the probabilities to create a probability mass function $f_{\overline{X}}(\overline{x})$ for $\overline{X}$. This is the sampling distribution for $\overline{X}$.

$\overline{x}$	1.00	1.50	2.00	2.50	3.00
$f_{\overline{X}}(\overline{x})$	0.25	0.30	0.29	0.12	0.04

Similarly we can create a probability mass function $f_{S^2}(s^2)$ for the sample variance $\overline{S}^2$. This is the sampling distribution for S^2. Note that we are treating the sample variance S^2 as a random variable now, because it's value depends on the particular sample that is gathered.

s^2	0.00	0.50	2.00
$f_{S^2}(s^2)$	0.38	0.42	0.20

Here is an R code that you can use to generate such sampling distributions:

```
################################
## EDIT the parameters below
################################

# list possible x values (population)
# it's best to keep this between 2 and 5 total values
x=c(1,2,3)

# weights for each x
# (number of tickets in the box)
w=c(5,3,2)

# sample size,
# use n=2 up to 6
# beyond that it may
# take up too much computer memory
n=2

################################
## DO NOT edit below here (without risk!)
################################

p=w/sum(w) # turn weights into probabilities
mu=sum(x*p) # population mean
sigsq=sum(x^2*p)-mu^2 # population variance

# now we create the list of all samples of size n,
# calculate sample statistics (mean and varaince of each sample)
S=expand.grid(replicate(n,x,simplify=FALSE))
pr=expand.grid(replicate(n,p,simplify=FALSE))

prob=cbind(0*1:length(S[,1])+1)

# calculate all sample means and their probabilities
xbar=0*prob
for (j in 1:n){
    prob=prob*pr[,j]
    xbar=xbar+S[,j]
}

# append table with probabilities and sample means
S$prob=prob
S$xbar=xbar/n

# calculate all sample varainces
```

```
var=0*prob
for (j in 1:n){
    var=var+(S[,j]-S$xbar)^2
}

# append table with probabilities and sample variances
S$var=var/(n-1)

# construct sampling distributions
# for sample mean and sample variance
xbar_vals=as.numeric(names(table(S$xbar)))
var_vals=as.numeric(names(table(S$var)))

xbar_probs=0*1:length(xbar_vals)
for (k in 1:length(xbar_vals)){
    xbar_probs[k]=sum(S$prob[S$xbar==xbar_vals[k]])
}

var_probs=0*1:length(var_vals)
for (k in 1:length(var_vals)){
    var_probs[k]=sum(S$prob[S$var==var_vals[k]])
}

# construct sampling distributions
xbar_samp_distr=rbind(xbar_vals, xbar_probs)
var_samp_distr=rbind(var_vals, var_probs)

xbar_mean=sum(xbar_samp_distr[1,]*xbar_samp_distr[2,])
xbar_var=sum(xbar_samp_distr[1,]^2*xbar_samp_distr[2,])-xbar_mean^2

var_mean=sum(var_samp_distr[1,]*var_samp_distr[2,])
var_var=sum(var_samp_distr[1,]^2*var_samp_distr[2,])-var_mean^2

# plot resulting sampling distributions
par(mfrow=c(2,1))
barplot(xbar_samp_distr[2,],
    names.arg=as.character(xbar_samp_distr[1,]),
    main="sample mean sampling distribution")
barplot(var_samp_distr[2,],
    names.arg=as.character(var_samp_distr[1,]),
    main="sample variance sampling distribution")

print(xbar_samp_distr)
print(var_samp_distr)
```