THE UNIVERSITY OF SYDNEY
SCHOOL OF MATHEMATICS AND STATISTICS

Tutorial Weeks 7 and 8

MATH1905: Statistics (Advanced) Semester 2, 2017

Web Page: `http://sydney.edu.au/science/maths/MATH1905`
Lecturer: Michael Stewart

There is a quiz in week 7 but these exercises for weeks 7 and 8 are provided ahead of time. Also please complete any unfinished exercises from week 6 and discuss any difficulties with your tutor, or attend a consultation session.

1. (**Multiple Choice**) The expected value, $E(X)$ of the random variable X having probability distribution

x	2	4	6
$P(X = x)$	0.1	0.3	0.6

is

(a) 4 (b) 0.3 (c) 0.5 (d) 5

2. Use R to simulate a set of $n = 25$ observations from the distribution given in the previous question using the `sample()` function, then find the sample mean:

```
x = c(2,4,6)                 # the possible outcomes

p = c(0.1,0.3,0.6)           # the probabilities associated
                             # with each outcome

samp = sample(x,size=25,replace=TRUE,prob=p)
mean(samp)
```

The mean you obtained is probably pretty close to the true value. Let's try running this experiment 10,000 times, each time generating a sample of size $n = 25$ and capturing the sample mean:

```
mx = 0                       # initialise an object that we
                             # will use to store the means

for(i in 1:10000){           # start the loop

                             # at each iteration generate
                             # a sample of size 25:
        samp = sample(x,size=25,replace=TRUE,prob=p)

                             # at each iteration calculate
                             # the sample mean and store it:
        mx[i] = mean(samp)
}
```

Now calculate the mean of the vector of means, `mx`. Is it closer to the true value? Also plot a box-plot, histogram and also an estimated density (like a smoothed histogram) using `plot(density(mx))`. What do you notice? Try increasing the sample size to $n = 500$. Does anything change? What phenomenon (or theorem) are we observing here?

3. (**Multiple Choice**) $X_1, X_2, \ldots, X_{25}$ represents a random sample from a distribution with mean $\mu = 10$ and standard deviation $\sigma = 20$. Indicate which of the following distributions is a good approximation to the distribution of $\bar{X} = \frac{1}{25}\sum_{i=1}^{25} X_i$.

(a) $N(10, 20^2)$
(b) $N(0, 1)$
(c) $N(10, 0.8)$
(d) $N(10, 4)$
(e) $N\left(10, \frac{20^2}{25}\right)$

4. X is binomial with $n = 100$ and $p = 0.4$ Which of the following normal distributions is a good approximation to the distribution of X?

(a) $N(100, 0.4)$
(b) $N(0, 1)$
(c) $N(40, 0.16)$
(d) $N(40, 24)$
(e) $N(40, 0.24)$

5. If $X \sim B(64, 0.5)$, the approximating normal variable Y is $N(32, 4^2)$. Using the correction for continuity, $P(32 < X < 36)$ is approximated by

(a) $P(32 < Y < 36)$
(b) $P(32.5 < Y < 35.5)$
(c) $P(32.5 < Y < 36.5)$
(d) $P(31.5 < Y < 36.5)$
(e) $P(31.5 < Y < 35.5)$

6. (a) Use R to find c if

(i) $P(t_{12} > c) = 0.01$ (ii) $P(t_5 \leq c) = 0.95$ (iii) $P(|t_{25}| > c) = 0.05$.

(b) Use R to find

(i) $P(t_{11} > 2.5)$ (ii) $P(|t_{15}| > 2.2)$.

7. (Illustration of the Central Limit Theorem using R)

(a) Use a loop (as below) to generate 1000 samples of size 25 from an exponential distribution with parameter $\lambda = 1$. For each sample compute the observed sample mean and store it in a vector called `sample.mean.obs` (as below).

```
sample.mean.obs=0                 # a place to store the result

for (i in 1:1000){                # begin for-loop

  x.obs=rexp(25,1)                # generate 25 "random"
                                  # exp(1) values

  sample.mean.obs[i]=mean(x.obs)  # compute the sample mean
                                  # for the ith sample

}                                 # end for-loop
```

Type `par(mfrow=c(2,2))` to prepare the graphics window for 4 plots. Note that at the end of the loop the object `x.obs` contains the 1000th (simulated) exponential sample.

Add the appropriate code to obtain a boxplot and histogram of the final (simulated) sample `x.obs` and then also the vector of sample means `sample.mean.obs` (you can add a useful heading by using a command like

```
boxplot(sample.mean.obs,horizontal=T,
        main="Boxplot of sample means (n=25)")
```

Comment on the shapes of both distributions. Repeat this question with 1000 samples of size 250.