
STAT 35000
Introduction to Statistics

Bonus 3 Project

Task 1:

- a) User define probability mass function. Suppose you have a random variable that can take the integer values 1,...10.

➤ `XX <- 1:10`

- b) Assign the pmf $p(x)=Pr(X=x)$ as: $p(1)=0.01$, $p(2)=1.2....$

➤ `PP <- c(.01, .12, .13, .14, .2, .2, .1, .05, .04, .01)`

- c) Check that this $p(x)$ sums to one

➤ `sum(PP)`

- d) The following puts two graphs one on top of the other

➤ `require(graphics)`

➤ `par(mfrow = c(2, 1))`

- e) Now plot the pmf you defined above (the specified option “h” means that the plot is vertical lines

➤ `plot(XX, PP, type="h", col=2, main="Pmf list", xlab="x", ylab="p(x) ")`

➤ `points(XX, PP, col=2); abline(h=0, col=3)`

- f) Obtain the cumulative probability distribution function (cdf)

➤ `QQ <- cumsum(PP)`

- g) Print all these values together

➤ `c(XX, PP, QQ)`

- h) Now plot the cdf, the option “s” means that the plot should be a staircase type. The $c(0, XX)$ and $c(0, QQ)$ adds a zero starting point.

➤ `plot(c(1, XX), c(0, QQ), type="s", ylab="F(x) ", col=2, xlab="x",
main="Cdf for user defined dist.")`

➤ `abline(h=0:1, col=4)`

Task 2:

Use the above approach to calculate and plot the pmf and the cdf for the first 11 values ($x=0, 1, 2, \dots, 10$) of the:

a) Binomial distribution with $n=10$ and $p=0.6$

```
➤ pbinom(XX, size =10, prob =0.6) # note that n stands for 'size'
➤ dbinom(XX, size =10, prob =0.6) # note that n stands for 'size'
```

b) and of the Poisson distribution with $\lambda=6$ (corresponding to $n=10$ and $p=0.6$)

```
➤ ppois(XX, lambda = 6)
➤ dpois(XX, lambda = 6)
```

Task 3: --

The purpose of this task is to introduce you to the term ***Sampling Distribution*** of a statistic. We consider a random sample of n independent $x_1, x_2, \dots, x_n$ observations drawn from a large population with (population) mean $E(X) = \mu$ and (population) variance $V(X) = \sigma^2$.

Usually, these two (population) parameters μ and σ^2 are unknown to us and we use the n observations in the sample to ***estimate*** these parameters. A natural ***estimator*** for the population mean μ is the sample mean

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

and a natural ***estimator*** of the population variance σ^2 is the sample variance

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

However, since the values of these two statistics $\bar{x}_n$, and S_n^2 depend upon the observed in values in the sample, these sample statistics are actually, also random variables. The distributions of their values (of such sample statistics) are called ***Sampling Distributions***.

A. You will now use R to generate (simulate), the sampling distribution of the statistic $\bar{x}_n$ (the sample average) when sampling from a distribution called the *Exponential distribution*.

a) Generate $M = 1000$ different random samples, each with $n = 16$ observations, from the

Exponential distribution, $X \sim \text{Exp}(\lambda = 0.1)$, which has mean $\mu = 10$ and a standard deviation $\sigma = 10$);

```
➤ M<-1000
➤ n<-16
➤ XSamples<-matrix(rep(0,n*M), M, n)
```

b) Generate a different random sample in each row of the matrix

```
➤ for(j in 1:M){XSamples[j , ]<- rexp(n, lambda=0.1)}
```

c) Calculate the sample mean of each sample (row)

```
➤ Xbars<-apply(XSamples, 1, mean)
```

d) Obtain the **sampling distribution** of $\bar{x}_n$, as the histogram of all these $M=1000$ sample averages $\bar{x}_n$ you simulated in the previous step.

```
➤ hist(Xbars, nclass=30, freq=F, main="Sampling Distribution
of Xbar when n=16")
```

e) From this histogram you can see that this sampling distribution is relatively symmetric, almost ‘bell-shaped’. Add ‘an approximated’ density curve to the histogram above.

```
➤ dens<-density(Xbars)
➤ lines(dens$x, dens$y, col=2)
```

f) The question however is what are the mean and the standard deviation of this (sampling) distribution? Get some summary statistics about this distribution;

```
➤ mean(Xbars)
➤ var(Xbars)
➤ sd(Xbars)
➤ summary(Xbars)
```

g) Let us use denote by $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ the mean and the standard deviation of the sampling distribution of $\bar{x}_n$, by part e) above, for a sample of size $n=16$;

$$\mu_{\bar{x}} \approx \underline{\hspace{2cm}} \quad \text{and} \quad \sigma_{\bar{x}} \approx \underline{\hspace{2cm}}$$

h) Repeat a)-g) above for $n=32, 64, 128$, and $\lambda=0.05, 0.1, 0.2, 0.5, 1$, and summarize the result in the table below.

Sampling Distribution of $\bar{x}_n$ when sampling from the Exponential Distribution $\text{Exp}(\lambda)$

$\lambda =$	$\mu =$	$\sigma =$	$n =$	$\mu_{\bar{x}} =$	$\sigma_{\bar{x}} =$
0.05	20	20	32		
			64		
			128		
0.1	10	10	32		
			64		
			128		
0.5	2	2	32		
			64		
			128		
1	1	1	32		
			64		
			128		

- B. Repeat all the steps in 1) above, but now simulate the sampling distribution of the statistic $\bar{x}_n$ when sampling from the Normal $N(\mu, \sigma^2)$ distribution. You need only to replace `rexp(n, lambda=??)` with `rnorm(n, mean=??, sd=??)` in step 1-b) above.

Sampling Distribution of $\bar{x}_n$ when sampling from the Normal Distribution $N(\mu, \sigma^2)$

$\mu =$	$\sigma =$	$n =$	$\mu_{\bar{x}} =$	$\sigma_{\bar{x}} =$
10	1	16		
		25		
		36		
10	3	16		
		25		
		36		
20	1	16		
		25		
		36		
20	3	16		
		25		
		36		