

**Solution:**

**a)**  Treating as the linear combination

=

From central limit Theoerm the properties of linear combination of random variables , has mean

Therefore

E() = 5

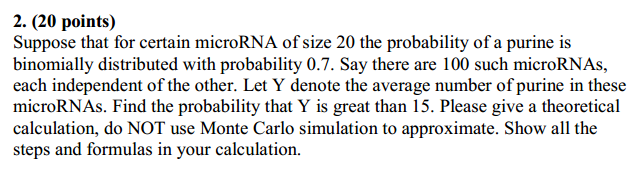
The variance has property

)22 = = = 1.8

Therefore, Var(1.8

Since standars deviation = sd() = =

**b) No,** I don’t think its possible to find P(2 < 5.1) as the sample size n = 5 for for  to cover distribution .



Solution: As each microRNA follows a binomial distribution with the probability of the purine being 0.7 . We can take that as the probability of success p=0.7 . Each microRNA is of size 20 , so n=20. We can now calculate E(X) using binomial distribution E(X) = np and variance Var(X) = np(1-p)

Substituting the values in the respective equation v get :

E(X) = 20 \* 0.7 = 14

Var(X) = 20 \* 0.7 \*(1 – 0.7) = 4.2

Therefore, the standard deviation is sd() = = 2.04939

When each microRNA is treated as an independent random variable Xi, we get {X1,…….,X100}. Y is the average number of purines in microRNA, we can write this formula as



{X1,…….,X100} are independent random samples drawn from a distribution with mean = 14 and sd = 2.04939. Here we can use CLT to calculate probability of calculate probability of events about the sample mean Y ,

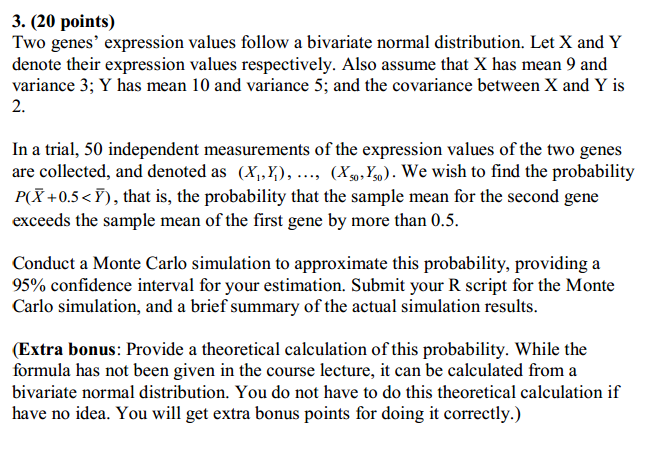
P( Y > 15 ) = 1 – P( Y ≤ 15 )

Therefore, P( Y > 15 ) can be approximately calculated in R as:

> 1-pnorm(15, mean=14,sd=2.04393/sqrt(100))

[1] 4.977259e-07

Therefore, P( Y > 15 ) 0. This is the probability that the average number of purines in the microRNA is greater than 15.



**Solution:** As the two random variables X and Y are not independent, we can model them via Monte Carlo simulation in R using the mvtnorm package. The code can be summarized as:

For this code we need to install mvtnorm so that we can use rmvnorm to generate realizations of X and Y.

We find the rvnorm , then the mean and variance for both x and y

Finally v find rnorm

rm(list=ls())

require(mvtnorm)

data<-rmvnorm(50,mean=c(9,10),sigma=matrix(c(3,2,2,5),nrow=2))

meanxy<-apply(data,1,mean)

varxy<-apply(data,1,var)

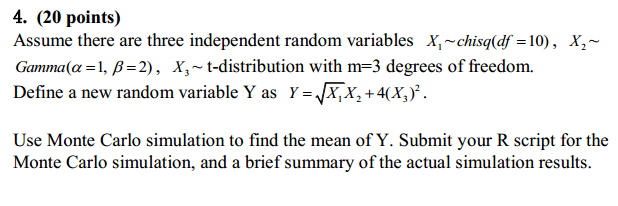
mean(meanxy) + c(-1,1)\*1.96\*sqrt(var(meanxy)/50) #95%CI for mean

mean(varxy) + c(-1,1)\*1.96\*sqrt(var(varxy)/50) #95%CI for variance

sqrt(1.912846)

sqrt(4.173590)

rnorm(10000, mean=c(9.27048, 10.40593), sd=c(1.383057, 4.173590))



Solution:

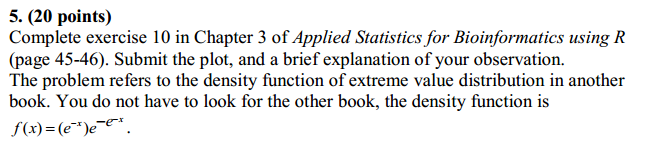
The random sampling of the distribution for  was modeled using rchisq. The random sampling of the distributioin for was modeled using rgamma and  was modeled using rt .

The mean for variable Y as  can be computed as

|  |
| --- |
| > rm(list=ls())  > #chi-square distribution  > X1 <- rchisq(10000, df=10)  > # gamma distribution  > X2 <- rgamma(10000, shape=1, scale=2)  > # t distribution  > X3 <- rt(10000, df=3)  > # calculate mean E(Y)  > Y <- sqrt(X1)\*(X2)+4\*(X3^2)  > meanY<- mean(Y)  > # Print mean E(Y)  > print (meanY)  [1] 18.27678 |
|  |
| |  | | --- | |  | |

The computed sampling to find the mean of Y was:

Mean(Y) = 18.27678



Solution: The density function is 

Taking 1000 random numbers from normal distribution N(mean=0,sd= 1) using rnorm , selecting maxima and performing it 1000 times to get 1000 maxima from normal distribution.

We then use the functions

an <- sqrt(2\*log(n)) - 0.5\*(log(log(n))+log(4\*pi))\*(2\*log(n))^(-1/2)

bn <- (2\*log(n))^(-1/2)

to get the set of maxim M = ( M – a )/b

Now we plot the density from the normalized maxima and density of normal distribution to get a histogram.

