

# STA5172 Random Variables in R

Course Instructor: Dr. Debdeep Pati

January 24, 2013

## 1 Counting in R

1. In R, combinations can be calculated using the choose () function. The choose function calculates the combinatorial of the parameters it is given. Choose (n, r), with parameter n and r, calculates  $\binom{n}{r}$ . For example:

```
> #calculate number of combinations of  
> #choosing 3 nucleotides from 4  
> choose(4,3)  
[1] 4
```

2. Gamma functions can be used together to perform any permutation or combinatorial counting procedure. For example, to calculate the numbers of unique 8-mer peptide arrangements taken from the 20 amino acids (order is important here so it is a permutation), simply use the formulae

```
> gamma(21)/gamma(13)  
[1] 5079110400
```

## 2 Random variables

For example, consider the RNA sequence: AUGCUUCGAAUGCUGUAUGAUGUC In this sequence there are 5 As, 9 Us, 6 Gs, and 4 Cs with a total of 24 residues. To model this sequence, the random variable X can be used where X represents the nucleotide residues. Because there are advantages to working with quantitative information, when the data is described qualitatively a random variable is used to assign a number to the non-numerical outcomes. For this experiment lets assign the random variable values representing A as 0, C as 1, G as 2 and U as 3. A small letter represents the outcome of the random variable, so little x can be used here. So, in probability terms, the model represented using the random variable X for this experiment is given in Table 6-1.

In R a simple histogram can be used to model the probability distribution function for this example.

```
> X<-c(0,1,2,3)  
> Prob<-c(0.208,0.167,0.25,0.375)  
> N<-c (A, C, G, U)  
> barplot(Prob,names=N,ylab="Probability", main="RNA Residue Analysis")
```

To model cdf

```
> CumProb<-c(0.208, 0.375, 0.625, 1)
> plot(X,CumProb,xlim=range(0,1,2,3,4), main="RNA Residue Analysis CDF",
xlab="X=", type="S")
```

### 3 Random variables

- Functions are provided to evaluate the cumulative distribution function  $P(X \leq x)$ , the probability density function and the quantile function (given  $q$ , the smallest  $x$  such that  $P(X \leq x) > q$ ), and to simulate from the distribution.

Distribution	R name	additional arguments
beta	beta	shape1,shape2, ncp
binomial	binom	size, prob
Cauchy	cauchy	location, scale
chi-squared	chisq	df, ncp
exponential	exp	rate
F	f	df1, df2, ncp
gamma	gamma	shape, scale
geometric	geom	prob
hypergeometric	hyper	m,n,k
log-normal	lnorm	meanlog, sdlog
logistic	logis	location, scale
neg. binomial	nbinom	size, prob
normal	norm	mean, sd
Poisson	pois	lambda
Students t	t	df, ncp
uniform	unif	min, max
Weibull	weibull	shape, scale
Wilcoxon	wilcox	m, n

- Plot the pmf and cdf of Binomial random variables

```
> x<-0:10
> y<-dbinom(0:10,10,0.16)
> data.frame("Prob"=y,row.names=x)
> plot(0:10,dbinom(0:10,10,0.16),,type='h',xlab="",ylab="Probability",
sub="Number of kids with blue eyes")
> par(mfrow=c(2,2))
> plot(0:10,dbinom(0:10,10,0.05),type='h',xlab="",ylab="Prob", sub="p=0.05")
> plot(0:10,dbinom(0:10,10,0.2),type='h',xlab="",ylab="Prob", sub="p=0.2")
> plot(0:10,dbinom(0:10,10,0.5),type='h',xlab="",ylab="Prob", sub="p=0.5")
> plot(0:10,dbinom(0:10,10,0.8),type='h',xlab="",ylab="Prob", sub="p=0.8")
> par(mfrow=c(2,2))
> plot(0:10,pbinom(0:10,10,0.05),type='s',xlab="",ylab="Prob", sub="p=0.05")
> plot(0:10,pbinom(0:10,10,0.2),type='s',xlab="",ylab="Prob",sub="p=0.2")
```

```
> plot(0:10,pbinom(0:10,10,0.5),type='s',xlab="",ylab="Prob",sub="p=0.5")  
> plot(0:10,pbinom(0:10,10,0.8),type='s',xlab="",ylab="Prob",sub="p=0.8")
```