# *R* computing for Business Data Analytics

Instructor: Dr. Howard Hao-Chun Chuang
Assistant Professor, Department of Management Information Systems
National Chengchi University, Taipei, Taiwan 116
chuang@nccu.edu.tw

Last Revised: August 2014

## 4.1 Discrete random variable

- Random variable

A random variable is a mapping/function from the sample space *S* to the real number line. A random variable *X* assigns a numerical value *x* of each possible outcome of an experiment.

Example: *X*=Toss a coin: *S*={head, tail}. Call a head 1 and a tail 0 => *S*={1, 0}.

A random variable is said to be *discrete* if its values assume integer points on the real number line. In other words, the outcomes are countably finite/infinite.

*Discrete distributions*: The set of all possible values of a discrete random variable along with their corresponding probabilities is called a discrete probability distribution.

*Probability density function* (*pdf*): For a discrete random variable *X*, the probability $P(X=x)$ is denoted by $p(x)$ and called the probability density (or mass) function. It satisfies:

$P(X=x) \geq 0$ or equivalently $p(x) \geq 0$

$\sum_x P(X=x) = 1$ or equivalently $\sum_x p(x) = 1$

$0 \leq P(X=x) \leq 1$ or equivalently $0 \leq p(x) \leq 1$

Example: Write out the distribution of tossing a fair coin.

Example: Let *X* be the face number after rolling a fair die. Then *x*=1, 2, 3, 4, 5, 6, and $P(X=1) = P(X=2)=P(X=3)=P(X=4)=P(X=5)=P(X=6)=1/6$. How does the distribution look like?

In the above two experiments, the probabilities of all possible outcomets/realizations are known prior to carrying out the experiment. This is not always the case. Sometimes we may have to adopt a relative frequency approach (using observed outcomes, *aka* data.).

Example: To obtain a model for campus Internet security in NCCU, the number of attacks occurring each week was observed over a period of 1 year. It was found that

0 attacks occurred in each of 9 weeks; 1 attack occurred in each of 14 weeks

2 attacks occurred in each of 13 weeks; 3 attacks occurred in each of 9 weeks

4 attacks s occurred in each of 4 weeks; 5 attacks occurred in each of 2 weeks

6 attacks occurred in each of 1 weeks

By obtaining the relative frequency of attacks, we can estimate the distribution using *R*

```
> attacks=c(9, 14, 13, 9, 4, 2, 1)
> probability=attacks/sum(attacks)
> probdens=round(probability, 2)
```

Write out the approximated distribution.

Visualize the *pdf*

```
> Attacks=0:6
> plot(Attacks, probdens, xlab= "# of attacks per week", ylab= "p(x)", type= "h")
```

We estimate probabilities from the *observed frequencies* of attacks over a year. Our hope is that the estimated probabilities are not too far away from the *true (but unknown)* probabilities. The greater the number of observations, the more accurate the estimated probabilities are.

*Cumulative distribution function* (*cdf*): The cumulative distribution function, $F(x)$, of a discrete random variable $X$ is defined by

$$F(x) = \sum_{k \le x} p(k) = P(X \le x)$$

We actually can show that

$$F(x) - F(x-1) = p(x) (\text{i.e., } P(X = x))$$

For the aforementioned number of attacks, write out the cdf.

Visualize the *cdf*

```
> cumprob=cumsum(probdens)
> plot(Attacks, cumprob, xlab= "# of attacks", ylab= "probability", type= "S")
```

- Benford's law

It was long assumed that the first digits of serial numbers ranged from 1 to 9 with equal likelihood (exercise – write a uniform *pdf*). Back in 1930s Benford showed that the number 1 occurs more frequently as a first digit than all other numbers. After examining diverse data sets, he found that the number 1 appeared as a first digit approximately 30% of the time in all cases. Hill (196) proposed a function that approximates the first digit frequencies.

$$f(x) = \log_{10}(1+\frac{1}{x}),\ x = 1, 2, \ldots, 9$$

Let's check this function in *R*.

```
> x=1:9
> prob=round(log10(1+1/x), 3)
> plot(x, prob, type="h")
```

Benford's law is widely applied to the detection of fraud given that fake data do not usually obey the first digit law. I hope this example helps you see that finding the correct distribution for a variable is a subtle and useful task. If well identified, the probability distribution can be applied to many types of real problems in our daily life.

- Summarizing random variables

The *expected value or mean* of a discrete random variable is defined as the weighted average of all possible values. The weights are the probabilities $p(x)$ ($P(X=x)$) of respective values $x$

$$\mu = E(X) = \sum_x xp(x)$$

A more generic form is

$$E(g(X)) = \sum_x g(x)p(x)$$

Now go back to the number of attacks discussed earlier, compute the expected value

Check the calculation in *R*.

```
> x=0:6
> probability=c(9/52, 14/52, 13/52, 9/52, 4/52, 2/52, 1/52)
> sum(x*probability)
```

In addition to the central tendency, the spread of a random variable is of great interest too. The most commonly used measure of spread is called *variance*

$$V(X) = \sigma^2 = E(X-\mu)^2 = \sum_x (x-\mu)^2 p(x)$$

The square root of the variance, $\sigma$, is called standard deviation.

Return to the number of attacks discussed earlier. Let's calculate the variance in *R*.

```
> x=0:6
> probability=c(9/52, 14/52, 13/52, 9/52, 4/52, 2/52, 1/52)
> mean=sum(x*probability)
> sum((x-mean)^2*probability)
```

We can further simulate the mean and variance.

```
> attacks_100=sample(c(0:6), 100, replace=T, prob=probability)
> attacks_1000=sample(c(0:6), 1000, replace=T, prob=probability)
> attacks_10000=sample(c(0:6), 10000, replace=T, prob=probability)
```

Use *mean*( ) and *var*( ) in *R* to obtain the simulated moments and compare those with actual.

- Properties of mean and variance

Suppose the monetary loss due to the number of attacks $X$ can be estimated in $USD using the function: Loss=10$X$+200. Compute (a) the expected weekly loss (b) the variance of the expected weekly loss. As such, we need to know $E(10X+200)$ and $V(10X+200)$.

Below are some useful properties. For any arbitrary constant $c$:

$$E(X+c) = \sum_x (x+c)p(x) = \sum_x xp(x) + c\sum_x p(x) = E(X)+c$$

$$E(cX) = \sum_x cxp(x) = c\sum_x xp(x) = cE(X)$$

$$V(X) = E(X^2) - E(X)^2$$

$$V(X+c) = E(X+c-E(X+c))^2 = E(X-E(X))^2 = V(X)$$

$$V(cX) = E(cX-E(cX))^2 = E(cX-cE(X))^2 = E(c(X-E(X)))^2$$

$$= \sum_x (c(x-E(X)))^2 p(x) = c^2 \sum_x (x-E(X))^2 p(x) = cE(X-E(X))^2 = c^2 V(X)$$

Given those properties, you should be able to compute $E(10X+200)$ and $V(10X+200)$.

**4.2 Some discrete distributions**

*R* has numerous built-in functions for handling the most commonly encountered probability distributions. Suppose a discrete random variable $X$ is of type *dist* with parameters $\boldsymbol{\theta}$, then

ddist(x, $\boldsymbol{\theta}$,…) returns $p(x)=P(X=x)$

pdist(x, $\boldsymbol{\theta}$,…) returns $F(x)=P(X\le x)$

qdist($p$, $\boldsymbol{\theta}$) returns the smallest $q$ for which $F(q)=P(X\le q)\ge p$

rdist($n$, $\boldsymbol{\theta}$) returns a vector of $n$ pseudo-random numbers from the distribution *dist*

Below we go over some famous discrete distributions, some of which will be visited later on when we step into statistical analysis of discrete data.

- Discrete uniform distribution http://en.wikipedia.org/wiki/Uniform_distribution_(discrete)

A random variable $X$ has a uniform discrete distribution if

$$P(X=x_i)=\frac{1}{n}, i=1,2,\dots,n$$

A nice example of this distribution is the face value of die rolling.

```
> x=1:6
> prob=rep(1/6, 6)
> plot(x, prob, type='h', xlab="Die rolling", ylab="Probability", main="pdf")
```

- Geometric distribution

A geometric random variable $X$ can be represented as the number of trials to the first success in an experiment, in which each trial is independent and reaches a success with probability $p$. The pdf $p(x)$ represents the probability that the first success will occur in the $x_{\text{th}}$ trial.

$$P(X=x)=(1-p)^{x-1}p, x=1,2,\dots,\infty$$

A useful property to remember is the sum of the geometric series with ratio (1-$p$)<1

$$1+(1-p)+(1-p)^2+(1-p)^3+\cdots=\frac{1}{p}$$

Example: Products produced by a machine have a 3% defective rate. What is the probability that the first defective occurs in the fifth item inspected? What is the probability that the first defective occurs in the first five inspections?

```
> dgeom(x=4, prob=0.03)
> pgeom(4, 0.03)
```

The geometric distribution has a *memoryless* property known as the "Markov property"

If $X$ is a geometric random variable, then for each integer $n$

$$P(X = x + n \mid X > n) = P(X = x)$$

Example: Products are inspected until the first defective is found. The first 10 trials have been found to be free of defectives. What is the probability that the first defective will occur in the 15th trial?

$$\begin{aligned} & P(X = 15 \mid X > 10) \\ &= \frac{P((X = 15) \cap (X > 10))}{P(X > 10)} \\ &= \frac{P(X = 15)}{P(X > 10)} \\ &= \frac{(1-p)^{14} p}{(1-p)^{10}} \\ &= (1-p)^4 p \\ &= P(X = 5) \end{aligned}$$

Note: It makes NO difference to probability that we already had 10 trials free of defectives. This Markov (memoryless) property has important implications for *gamblers*. Suppose a gambler is betting on "heads" in the tossing of a fair coin. After 10 successive tosses of "tails", what is the probability of having a "head" in the next trial? Coins have NO memory. Unfortunately, gamblers tend to feel that given that heads and tails are 50:50, long sequences of tails are *increasingly likely* to be followed by heads. That is why people lose all. Beware!

- Bernoulli distribution

Suppose an experiment has a sample space that consists of two outcomes, a "success" with probability $p$ and a "failure" with probability 1-$p$, we can model the outcome as a Bernoulli random variable $X$ with the pdf

$$\begin{aligned} P(X = 0) &= 1 - p \\ P(X = 1) &= p \end{aligned} \quad \text{where } 0 \le p \le 1$$

An example of this distribution is the outcome of tossing a coin. The Bernoulli distribution is also referred to as an *indicator variable* and sets up the foundation of logistic regression that will be introduced later on. It is a special case of the binomial distribution below.

In $R$, you use the Bernoulli distribution using *dbinom*(*x, size=1, prob=p,…*).

- Binomial distribution

An experiment has $n$ independent trials, and each trial has a "success" with probability $p$ and a "failure" with probability 1-$p$. We can model *the number of successes* as a binomial random variable $X$ with its pdf $p(x)$

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x},\ x=0,1,2,...,n$$

The random variable gets its name from the *binomial expansion*:

$$\sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x}=(p+1-p)^n=1$$

We can also easily show that the Bernoulli distribution is a special case with $n$=1.

Example: 20% of integrated circuit chips on a production line are known to be defective. A sample of 20 chips is selected at regular intervals for inspection. Up to how many defectives $k$ will the sample contain at least 95% certainty? We want to find $k$ such that

$$P(X\le k)\ge 0.95$$

```
> qbinom(0.95, size=20, prob=0.2)
```

So, if chips from this production process are packed in batches of size 20, then 95% of these batches will contain less than or equal to $k$ defectives.

- Poisson distribution (check *help*(*dpois*) in *R*)

Poisson distribution is probably one of the famous discrete distributions. This distribution is used as a model for rare events, and event occurring at random over time and space. I cannot over-emphasize the importance of Poisson distribution as it has many important applications in inventory theory, queuing theory, etc. Its pdf $p(x)$ is

$$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!},\ x=0,1,2,...,\infty$$

Poisson distribution, in fact, is proven as the *limiting distribution* of the binomial distribution

Define $p=\dfrac{\lambda}{n}$, $\displaystyle\lim_{n\to\infty}\binom{n}{x}p^x(1-p)^{n-x}=\frac{\lambda^x e^{-\lambda}}{x!},\ x=0,1,2,...,\infty\ (\because \lim_{n\to\infty}(1+\frac{x}{n})^n=e^x)$

Here are two properties you should remember for a random variable $X$~Pois($\lambda$)

$E(X)=Var(X)=\lambda$ ($\lambda$ is the arrival rate of a Poisson process).

If $X$~Pois ($\lambda_1$) and $Y$~Pois ($\lambda_2$) are independently distributed, then $X+Y$~Pois($\lambda_1+\lambda_2$)

- Negative binomial distribution (check *help*(*dnbinom*) in *R*)

Let a random variable *X* be the number of failures before the $r_{th}$ success, in a sequence of independent Brenoulli($p$) trials, *X* is said to have a negative binomial distribution with pdf

$$P(X=x)=\binom{r+x-1}{r-1}p^r(1-p)^x,\ x=0,1,2,...,\infty$$

In fact, let $Y_1,\ldots,Y_r$ be *iid* geometric($p$) random variables, then

$$X=Y_1+\cdots Y_r\sim \text{nbinom}(r,p)$$

Moreover, we can prove that for $X$~Poisson($\lambda$) and $\lambda$ conforms to a gamma distribution, *X* turns out to be a negative binomial random variable. So, this distribution is actually a *mixed Poisson distribution*. Both the Poisson and negative binomial distributions play critical roles in *count data* analysis. We will revisit those two distributions as the semester proceeds.

Please keep it in mind that the list of discrete distributions above is by no means exhaustive. Nonetheless, lecture 4 equips you with necessary (although not sufficient) knowledge about discrete random variables, some of which, depending on how creative you are, can be applied to solve many practical problems and model socio-technical phenomena/observations.

**4.3 Continuous random variable**

If a random variable *X* takes all values in a finite or infinite interval, then *X* is said to be a continuous random variable.

Example: Let *X* denote the CPU time to process a program. Here, $0<x<\infty$.

*Probability density function*(*pdf*): For a continuous random variable *X*, its *pdf* $f(x)$ has the following properties:

$$\int_{-\infty}^{\infty} f(x)dx=1$$

$$\int_a^b f(x)dx=P(a\le X\le b)$$

$$f(x)\ge 0$$

Note that for a continuous variable $P(X=k)=0$ for all *k*. Here we have the probability density, which is the probability of hitting a small region around *k* divided by the size of the region.

*Cumulative distribution function*(*cdf*): The cdf of a continuous random variable *X* is usually denoted by $F(x)$ and defined as

$$F(x)=P(X<x)=P(X\le x)=\int_{-\infty}^{x} f(t)dt$$

Analogous to the discrete case, when $X$ is continuous, it has the following properties

$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)dx,$ or more generally, $E(g(X)) = \mu = \int_x g(x)f(x)dx$

$V(X) = \sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx,$ and invariantly, $V(X) = E(X^2) - E(X)^2$

Again, for any arbitrary constant $c$:

$E(X+c) = E(X)+c$

$E(cX) = cE(X)$

$V(X+c) = V(X)$

$V(cX) = c^2V(X)$

**4.4 Some continuous distributions**

*R* has numerous built-in functions for handling the most commonly encountered probability distributions. Suppose a continuous random variable $X$ is of type *dist* with parameters $\boldsymbol{\theta}$, then

ddist(x, $\boldsymbol{\theta}$,…) returns the density $f(x)$

pdist(x, $\boldsymbol{\theta}$,…) returns $F(x)=P(X\leq x)$

qdist($p$, $\boldsymbol{\theta}$) returns the smallest $q$ for which $F(q)=P(X\leq q)\geq p$

rdist($n$, $\boldsymbol{\theta}$) returns a vector of $n$ pseudo-random numbers from the distribution *dist*

Below we go over some continuous distributions, which any data analysts must be aware of.

- Uniform distribution (check *help*(*dunif*) in *R*)

If the probability that a random variable $X$ lies in a subinterval of $[a, b]$ depends only on the length of the subinterval and not on its location, then $X$ is said to have a uniform/rectangular distribution on $[a, b]$. Its *pdf* is given by

$$f(x) = \frac{1}{b-a}, \; a \leq x \leq b$$

The uniform distribution is **indispensable** to simulation as it enable us to generate pseudo-random numbers from all sorts of probability distributions. We will get to the bottom of it in the next lecture on *Monte-Carlo simulation*.

Now we shift gears to three *lifetime* models. Let $X \geq 0$ be the time until some event occurs. We can define the *survivor function* $G(x) = P(X > x) = 1 - F(x)$, which is the probability that the event will not happen until time $x$. Also define the *hazard function* $\lambda(x)$, which is the rate at which event occurs at time $x$, that is

$$\lambda(x)dx = P(X \in (x, x+dx] \mid X > x) = \frac{f(x)dx}{G(x)} \Rightarrow \lambda(x) = \frac{f(x)}{G(x)}$$

Keep the term hazard function/rate in mind for the moment. We will come back to it once we have opportunities to deal with time-to-event data. But do understand that $\lambda(x)$ is the key that makes differences in the three distributions to be introduced below.

- Exponential distribution (check *help*(*dexp*) in *R*)

If $\lambda(x) = \lambda$ (i.e., a constant rate), then $X$ has an exponential distribution and write $X$~exp($\lambda$).

$$f(x) = \lambda e^{-\lambda x},\ x > 0$$

This $\lambda$ is exactly the lambda you have seen in Poisson($\lambda$) in lecture 4. Mathematically we can show how the two distributions are intertwined through stochastic processes.

I hope you also remember the geometric distribution and its Markov/memoryless property (as well as its implications to gamblers who bet on coin-tossing). $X$~exp($\lambda$), being a continuous analogue of the geometric distribution, has this memeoryless property too (given the constant $\lambda(x)$). It is expressed as follows for $s, t \geq 0$

$$P(X > s + t \mid X > s) = P(X > t)$$

The proof of this property will be left as an exercise in your homework.

The exponential distribution has salient applications in queuing theory and actuarial science. However, we do not have time to take a deep dive into the model. Below is a simple example that illustrates the idea of the exponential distribution.

Example: If jobs arrive every 15 seconds on average, $\lambda$= 4 per minute. (a) What is the probability of waiting less than or equal to 30 seconds? (b) What is the probability that at least twenty minutes will elapse between accidents? (c) What is the maximum waiting time between two job submissions with 95% confidence?

(a) We know $P(X < x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$. So,

$$P(X \leq 0.5 \mid \lambda = 4) = \int_{0}^{0.5} \lambda e^{-\lambda 0.5} dt = \int_{0}^{0.5} 4e^{-4*0.5} dt = 0.86$$

In *R* it is just

>pexp(???, 4)

(b) We want to find $P(X \geq 20 \mid \lambda = 4) = \int_{20}^{\infty} \lambda e^{-\lambda 20} dt$

How to do this in *R*?

>

(c) We need to find $k$ such that

$$P(X \le k) = 0.95$$

>

- Gamma distribution (check *help*(*dgamma*) in *R*)

In lecture 4 we mention that if we sum independent geometric random variables, we get a negative binomial random variable. While the exponential distribution is the continuous analogue of the geometric distribution, the continuous analogue of the negative binomial distribution is the gamma distribution.

Let a random variable $X$ be the sum of $m$ independent exp($\lambda$) random variables. $X$ has the *pdf*

$$f(x) = \frac{1}{\Gamma(m)} \lambda^m x^{m-1} e^{-\lambda x} \text{ for } x \ge 0 \text{ and } m, \lambda > 0$$

Note that the definition holds for all $m \in (0, \infty]$, not just integer values. In the special case where $m$ is integer, the gamma distribution is known as the Erlang distribution. The gamma distribution is very flexible and capable of characterizing various sorts of data. Hence, it has great utility when dealing with skewed data that violates normality assumptions held in many statistical models.

*Exponential* and *gamma* distributions play critical roles in *survival analysis* and *event study*. They have numerous applications in social, management, natural, medical, and engineering sciences. We will revisit these distributions when we proceed to *generalized linear modeling*.

The last three models I want to talk about are sampling distributions, which are important in statistical data analysis when dealing with random samples. A sampling distribution is the distribution of a statistic such as $\bar{x}$ (http://en.wikipedia.org/wiki/Sampling_distribution).

- Normal (or Gaussian) distribution (check *help*(*dnorm*) in *R*)

A normally distributed random variable $X \sim N(\mu, \sigma^2)$ has the *pdf*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

Normal is SUPER popular, but DO NOT take Normal for granted and apply it to all data!

Let's generate some normal curves.

```
>curve(dnorm(x, 1.5, .5), from=0, to=3, ylab="f(x)")
```

Do you see the bell-shape and the symmetry?

Set two normal distributions with different means ($\mu$)

```
>x=seq(0.5, 3, len=101)
>y=cbind(dnorm(x, 1.65, .25), dnorm(x, 1.85, .25))
>matplot(x, y, type="l", xlab="", ylab="f(x)")
```

Set two normal distributions with different variances ($\sigma^2$)

```
>x=seq(4, 6, len=101)
>y=cbind(dnorm(x, 10, 1), dnorm(x, 10, 2))
>matplot(x, y, type="l", xlab="", ylab="f(x)")
```

Now you should have a sense of how the normal distribution behaves. A remarkable property is that if $X \sim N(\mu_1, \sigma_1^2)$ & $Y \sim N(\mu_2, \sigma_2^2)$ indepedent of $X$, then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Even if you cannot prove this mathematically, you can verify it experimentally. Here is how.

```
>z1=rnorm(10000, 1, 1)
>z1=rnorm(10000, 1, 2)
> z=z1+z2
>par(las=1)
>hist(z, breaks=seq(-10, 14, 0.2), freq=F)
>f=function(x) {exp(-(x-2)^2/10)/sqrt(10*pi)}
>x=seq(-10, 14, 0.1)
>lines(x, f(x))
```

We can see that the scaled histogram is very close to the theoretical density.

You may remember in your old (or just new like unused?) statistics text book, there is just one table for normal distribution. Here it comes: the case $Z = \frac{X - \mu}{\sigma} \sim N(0,1)$ is called the **standard normal**. All normal deviates can be transformed into *Z*, which has the pdf $\phi$ and the cdf $\Phi$. The functional form of $\phi$ is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

Below are some critical numbers to remember (verify these in *R* using *pnorm*( ))

$P(-1.645<Z<1.645)=0.90$; $P(-1.96<Z<1.96)=0.95$; $P(-2.58<Z<2.58)=0.99$

Given $Z=(X-\mu)/\sigma$, we know that $P(\mu-1.645\sigma<X<\mu+1.645\sigma)=0.90$, and so on and so forth…

Last, how do we find the *k* such that $P(X \le k \mid \mu, \sigma^2)$ equals to, for example, 0.90?

- $\chi^2$ (chi-squared) distribution (check *help*(*dchisq*) in *R*)

Suppose $Z_1, ..., Z_v$ are iid *N*(0,1) random variables, then $X = Z_1^2 + \cdots Z_v^2$ has a chi-squared distribution with degrees of freedom, and we write $X \sim \chi_v^2$. Its *pdf* is given by

$$f(x) = \frac{1}{2^{v/2}\Gamma(v/2)} x^{(v/2)-1} e^{-x/2}, \; x \ge 0$$

We can show that $X \sim \chi_v^2$ is the special case of the gamma distribution. The $\chi^2$ distribution is crucial for many statistical tests (e.g., goodness-of-fit test). Note that a random variable of the *F* distribution (in *ANOVA*) is just the ratio of two properly scaled $\chi^2$ random variables.

- Student's *t* distribution (check *help*(*dt*) in *R*)

If *X*~*N*(0, 1) and $Y \sim \chi_v^2$ independent of *X*, then

$$T = \frac{X}{\sqrt{Y/v}}$$

is said to have a *t* distribution with *v* degrees of freedom. $T \sim t_v$ has the density.

$$f(x) = \frac{\Gamma((v+1)/2)}{\sqrt{v\pi}\Gamma(v/2)}(1+\frac{x^2}{v})^{-(v+1)/2}, \; -\infty < x < \infty$$

You must have already seen the *t* distribution in two-sample *t*-tests for a difference in mean. The *t* distribution is fundamental for regression analysis as it is used to construct *confidence intervals* for the mean when the population variance is unknown.

We will revisit the sampling distributions as we move into parameter estimation, simulation, and data analysis.

All the distributions we have seen in lecture 4 are *univariate* ones. For the sake of time, I decide to skip *multivariate* distributions without compromising your learning. You should know that there are way too many distributions that we just have no time to cover them all (see http://cran.r-project.org/web/views/Distributions.html). There are plenty of others for you to explore when you encounter some unique data sets in the future. That said, those you have learnt from lecture 4 are important for the rest of semester. Now you are equipped with fundamental knowledge and ready to move into the next phase – statistical data analysis.