3.5. Normal Distributions

1. *Formulas from Calculus.*

$$\int_0^\infty \exp\left[-x^2\right] dx = \sqrt{\pi}/2. \tag{1}$$

$$\int_{-\infty}^\infty \exp\left[-x^2\right] dx = \sqrt{\pi}. \tag{2}$$

If we replace x with $x/\sqrt{2}$ in equation (2), we get:

$$\int_{-\infty}^\infty \exp\left[-x^2/2\right] dx = \sqrt{2\pi}. \tag{3}$$

2. If we divide equation (3) by $\sqrt{2\pi}$, we get

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] dx = 1. \tag{4}$$

Define the function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right], \qquad -\infty < x < \infty, \tag{5}$$

then $f(x) > 0$ and we can rewrite equation (4) as:

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

3. If summary, the function

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right], \quad -\infty < x < \infty, \tag{5}$$

qualifies as a pdf.

4. *Definition.* If X is a continuous random variable, and X has the pdf $f(x)$ (as defined in equation (5)), then we say X has the standard normal distribution.

5. *More Formulas.* Again, we have

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}. \tag{1}$$

If we replace x with $\sqrt{x}$ in equation (1), we get

$$\int_0^\infty e^{-x} d(\sqrt{x}) = \frac{\sqrt{\pi}}{2}.$$

It follows that

$$\int_0^\infty \frac{1}{\sqrt{x}} e^{-x} dx = \sqrt{\pi}, \quad \text{that is,} \quad \Gamma(1/2) = \sqrt{\pi}.$$

6. *Definition.* If X is a continuous random variable with pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty,$$

where $\sigma > 0$, then we say X has the normal distribution with parameters (μ, σ^2). The short notation for the normal distribution with parameters (μ, σ^2) is $N(\mu, \sigma^2)$.

7. *Formula.* If X has the $N(\mu, \sigma^2)$ distribution, then

$$E(X) = \mu, \quad Var(X) = \sigma^2.$$

and X has mgf:

$$M(t) = \exp(\mu t + \sigma^2 t^2/2).$$

8. For example, if $X \sim N(3, 5)$, then

$$E(X) = 3, \quad Var(X) = 5.$$

9. The $N(0,1)$ distribution is called the standard normal distribution.

10. *Theorem.* If $X \sim N(0,1)$, then its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right], \quad -\infty < x < \infty, \tag{5}$$

$$E(X) = 0, \quad Var(X) = 1.$$

and X has mgf:

$$M(t) = \exp(t + t^2/2).$$

11. *Theorem.* If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

— *The proof will be given in the next chapter.*

12. *Theorem.* Let X be a random variable. Suppose that $E(X) = \mu$, $Var(X) = \sigma^2 > 0$. If we define

$$Y = \frac{X - \mu}{\sigma},$$

then $E(Y) = 0$, $Var(Y) = 1$. The variable Y is called the standardization of X.

13. Now we can restate Theorem 11 as follows:

14. *Theorem.* $N(0,1)$ is the standardization of $N(\mu,\sigma^2)$.

15. *Theorem.* If X is $N(0,1)$, then $X^2 \sim \chi^2(1)$.

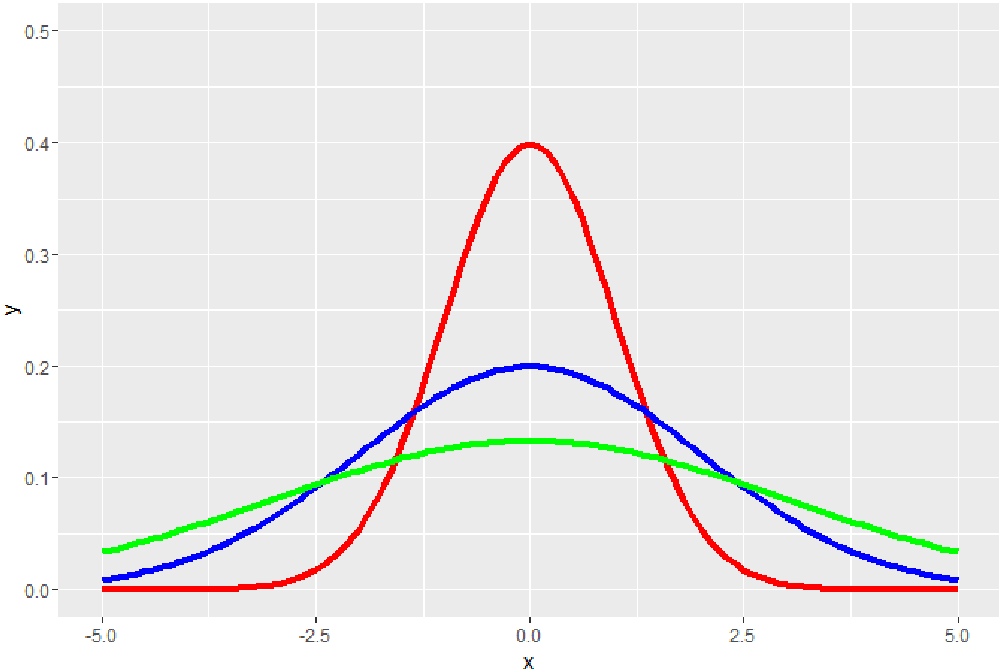
    — *Proof will be given later.*

16. On the next page, three curves are plotted:

    (a) The red curve is the pdf of $N(0,1)$ distribution;

    (b) The blue curve is the pdf of $N(0,4)$ distribution;

    (c) The green curve is the pdf of $N(0,9)$ distribution;

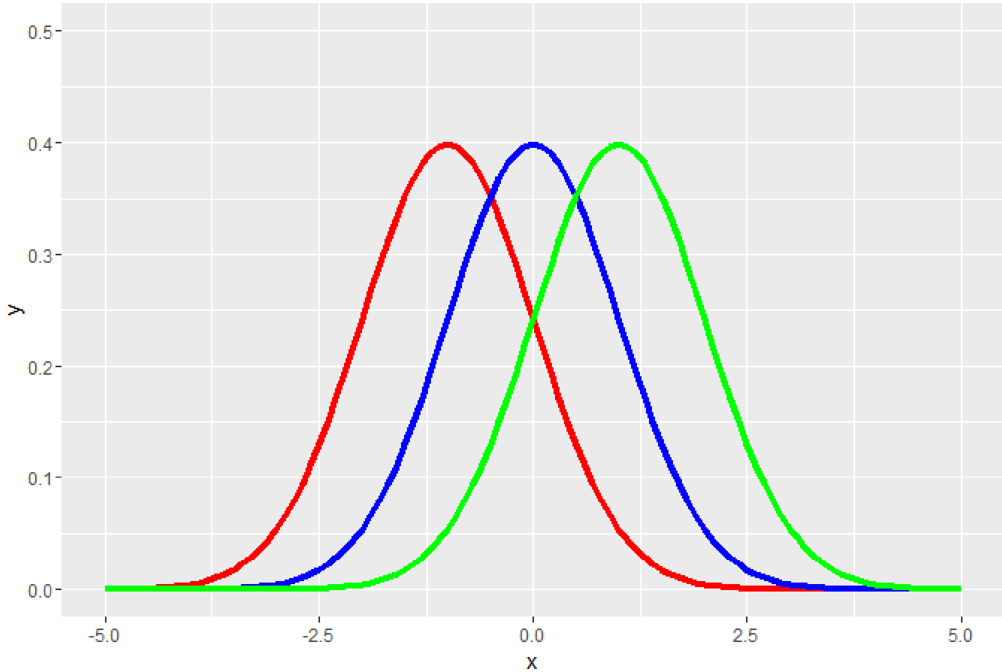

0.5
0.4
0.3
0.2
0.1
0.0
y
-5.0
-2.5
0.0
2.5
5.0
x

17. On the next page, three curves are plotted:

    (a) The red curve is the pdf of $N(-1, 1)$ distribution;

    (b) The blue curve is the pdf of $N(0, 1)$ distribution;

    (c) The green curve is the pdf of $N(1, 1)$ distribution.


0.5
0.4
0.3
0.2
0.1
0.0
y
-5.0
-2.5
0.0
2.5
5.0
x

18. *R code:*

```
library(ggplot2);
h <- ggplot(data.frame(x = c(-5, 5)), aes(x = x)) ;
h<-h+stat_function(fun=dnorm, geom = "line",size=1.5,col="red",
args=list(mean=0,sd=1));
h<-h+stat_function(fun=dnorm, geom = "line",size=1.5,col="blue",
args=list(mean=0,sd=2));
h<-h+stat_function(fun=dnorm, geom = "line",size=1.5,col="green",
args=list(mean=0,sd=3));
h <- h+ lims(y = c(0, 0.5));
h
```

19. *R code:*

```
library(ggplot2);
h <- ggplot(data.frame(x = c(-5, 5)), aes(x = x)) ;
h<-h+stat_function(fun=dnorm, geom = "line",size=1.5,col="red",
args=list(mean=-1,sd=1));
h<-h+stat_function(fun=dnorm, geom = "line",size=1.5,col="blue",
args=list(mean=0,sd=1));
h<-h+stat_function(fun=dnorm, geom = "line",size=1.5,col="green",
args=list(mean=1,sd=1));
h <- h+ lims(y = c(0, 0.5));
h
```