SCMA470 Risk Analysis and Credibility

Pairote Satiracoo

2021-08-26

Contents

Chapter 1

Basic Probability Concepts

1.1 Random Variables

Definition 1.1. *Let S be the sample space of an experiment. A real-valued function $X : S \to \mathbb{R}$ is called a* ***random variable*** *of the experiment if, for each interval $I \subset \mathbb{R}$, $\{s : X(s) \in I\}$ is an event.*

Random variables are often used for the calculation of the probabilities of events. The real-valued function $P(X \leq t)$ characterizes X, it tells us almost everything about X. This function is called the **cumulative distribution function** of X. The cumulative distribution function describes how the probabilities accumulate.

Definition 1.2. *If X is a random variable, then the function F defined on $\mathbb{R}$ by*

$$F(x) = P(X \leq x)$$

is called the ***cumulative distribution function*** *or simply* ***distribution function (c.d.f)*** *of X.*

Functions that define the probability measure for discrete and continuous random variables are the probability mass function and the probability density function.

Definition 1.3. *Suppose X is a discrete random variable. Then the function*

$$f(x) = P(X = x)$$

that is defined for each x in the range of X is called the ***probability mass function*** *(p.m.f) of a random variable X.*

Definition 1.4. *Suppose X is a continuous random variable with c.d.f F and there exists a nonnegative, integrable function f, $f : \mathbb{R} \to [0, \infty)$ such that*

$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$

Then the function f is called the ***probability density function*** *(p.d.f) of a random variable X.*

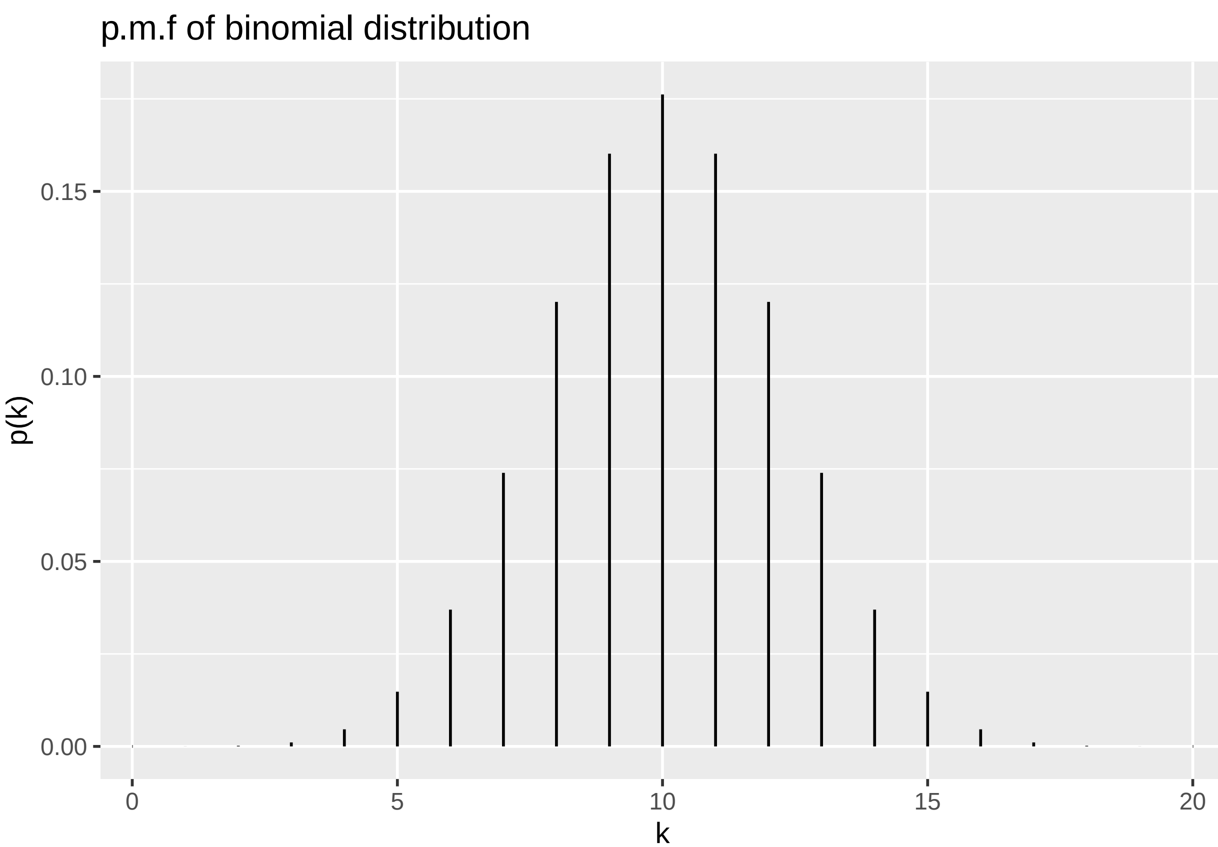
1.1.1 R Functions for Probability Distributions

In R, density, distribution function, for the Poisson distribution with parameter λ is shown as follows:

Distribution	Density function: $P(X = x)$	Distribution function: $P(X \leq x)$	Quantile function (inverse c.d.f.)	random generation
Poisson	`dpois(x, lambda, log = FALSE)`	`ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)`	`qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)`	`rpois(n, lambda)`

For the binomial distribution, these functions are pbinom, qbinom, dbinom, and rbinom. For the normal distribution, these functions are pnorm, qnorm, dnorm, and rnorm. And so forth.

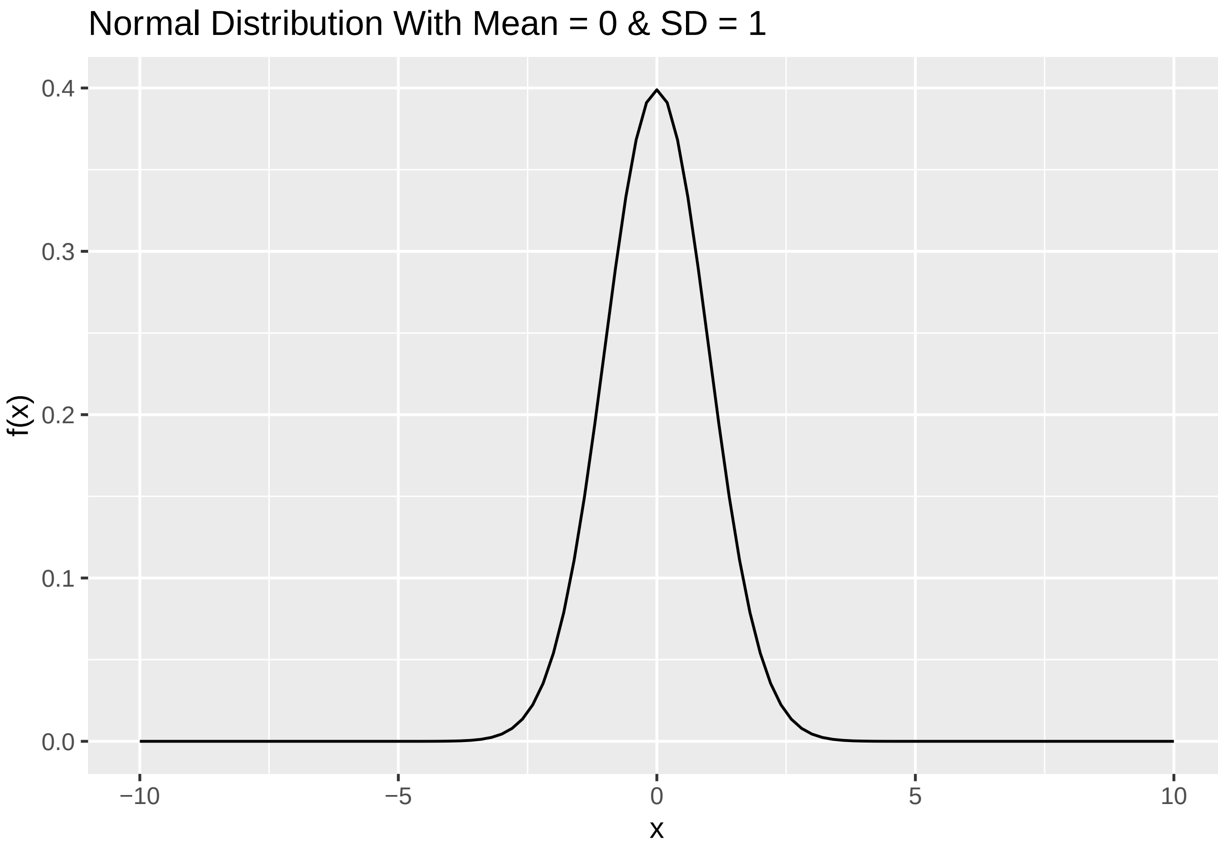
```
library(ggplot2)
x <- 0:20
myData <- data.frame( k = factor(x), pK = dbinom(x, 20, .5))
ggplot(myData,aes(k,ymin=0,ymax=pK)) +
  geom_linerange() + ylab("p(k)") +
  scale_x_discrete(breaks=seq(0,20,5)) +
  ggtitle("p.m.f of binomial distribution")
```



To plot continuous probability distribution in R, we use stat_function to add the density function as its arguement. To specify a different mean or standard deviation, we use the **args** parameter to supply new values.

```
library(ggplot2)
df <- data.frame(x=seq(-10,10,by=0.1))
ggplot(df) +
    stat_function(aes(x),fun=dnorm, args = list(mean = 0, sd = 1))  +
    labs(x = "x", y = "f(x)",
```

```
    title = "Normal Distribution With Mean = 0 & SD = 1")
```



1.2 Expectation

Definition 1.5. *The **expected value** of a discrete random variable X with the set of possible values A and probability mass function $f(x)$ is defined by*

$$\mathrm{E}(X) = \sum_{x \in A} x f(x)$$

The **expected value** of a random variable X is also called the mean, or the mathematical expectation, or simply the expectation of X. It is also occasionally denoted by $\mathrm{E}[X]$, μ_X, or μ.

Note that if each value x of X is weighted by $f(x) = P(X = x)$, then $\sum_{x \in A} x f(x)$ is nothing but the weighted average of X.

Theorem 1.1. Let X be a discrete random variable with set of possible values A and probability mass function $f(x)$, and let g be a real-valued function. Then $g(X)$ is a random variable with

$$\mathrm{E}[g(X)] = \sum_{x \in A} g(x) f(x)$$

Definition 1.6. *If X is a continuous random variable with probability density function f , the **expected value** of X is defined by*

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Theorem 1.2.

- *Let X be a continuous random variable with probability density function $f(x)$; then for any function $h : \mathbb{R} \to \mathbb{R}$,*

$$\mathrm{E}[h(X)] = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx$$

- *

Theorem 1.3. Let X be a random variable. Let $h_1, h_2, ..., h_n$ be real-valued functions, and $a_1, a_2, \dots, a_n$ be real numbers. Then

$$\mathrm{E}[a_1 h_1(X) + a_2 h_2(X) + \cdots + a_n h_n(X)] = a_1 \mathrm{E}[h_1(X)] + a_2 \mathrm{E}[h_2(X)] + ... + a_n \mathrm{E}[h_n(X)]$$

Moreover, if a and b are constants, then

$$\mathrm{E}(aX + b) = a\mathrm{E}(x) + b$$

1.3 Variances of Random Variables

Definition 1.7. *Let X be a discrete random variable with a set of possible values A, probability mass function $f(x)$, and $\mathrm{E}(X) = \mu$. then $\mathrm{Var}(X)$ and σ_X, called the* ***variance*** *and* ***standard deviation*** *of X, respectively, are defined by*

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mu)^2] = \sum_{x \in A} (x - \mu)^2 f(x),$$

$$\sigma_X = \sqrt{\mathrm{E}[(X - \mu)^2]}$$

Definition 1.8. *If X is a continuous random variable with $\mathrm{E}(X) = \mu$, then $\mathrm{Var}(X)$ and σ_X, called the* ***variance*** *and* ***standard deviation*** *of X, respectively, are defined by*

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2\, f(x)\, dx,$$

$$\sigma_X = \sqrt{\mathrm{E}[(X - \mu)^2]}$$

We have the following important relations

$$\mathrm{Var}(x) = \mathrm{E}(X^2) - (\mathrm{E}(x))^2,$$

$$\mathrm{Var}(aX + b) = a^2\ Var(X), \quad \sigma_{aX+b} = |a| \sigma_X$$

where a and b are constants.

1.4 Moments and Moment Generating Function

Definition 1.9. *For $r > 0$, the rth moment of X (the rth moment about the origin) is $\mathrm{E}[X^r]$, when it is defined. The rth central moment of a random variable X (the rth moment about the mean) is $\mathrm{E}[(X - \mathrm{E}[X])^r]$.*

Definition 1.10. *The skewness of X is defined to be the third central moment,*

$$\mathrm{E}[(X - \mathrm{E}[X])^3],$$

and the coefficient of skewness to be given by

$$\frac{\mathrm{E}[(X - \mathrm{E}[X])^3]}{(\mathrm{Var}[X])^{3/2}}.$$

Definition 1.11. *The coefficient of kurtosis of X is defined by*

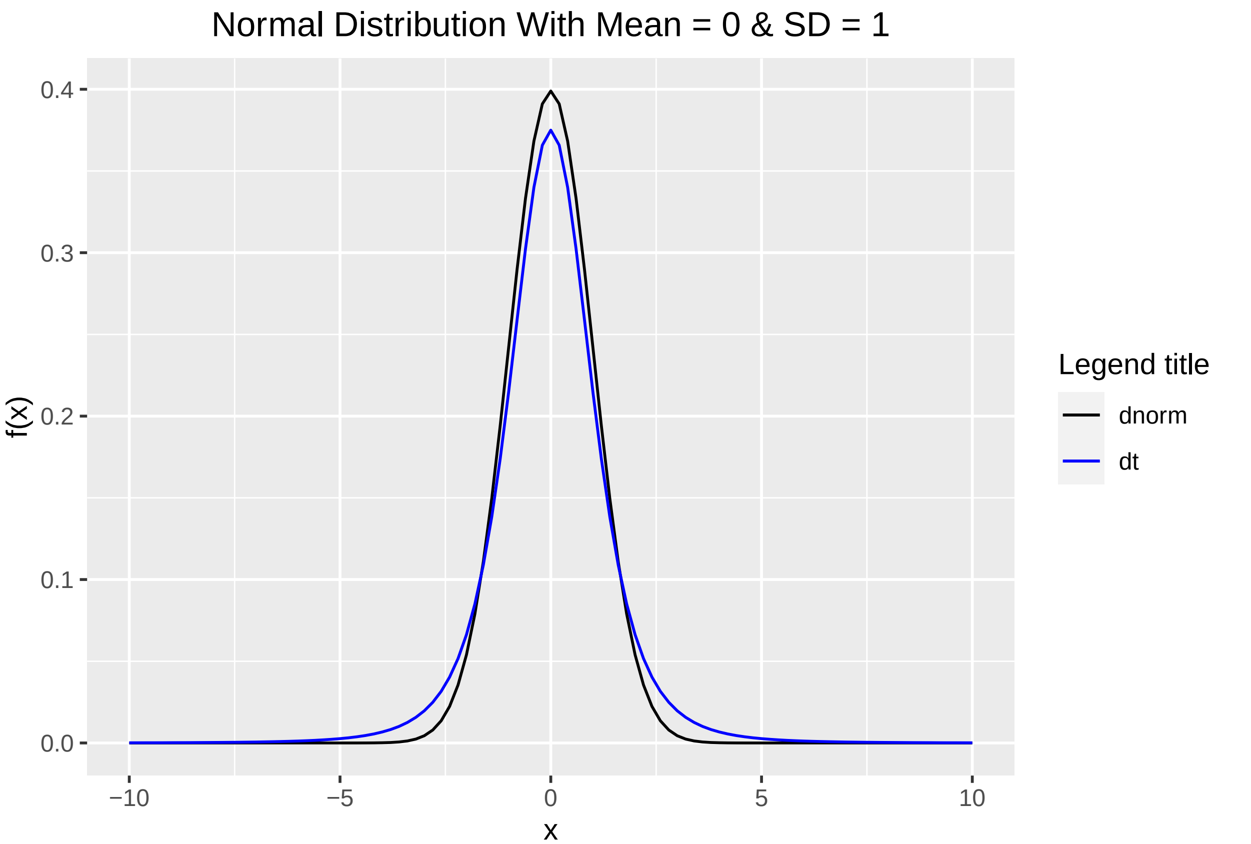
$$\frac{\mathrm{E}[(X - \mathrm{E}[X])^4]}{(\mathrm{Var}[X])^{4/2}}.$$

Note In the formula, subtract from the mean and normalise or divide by the standard deviation center and scale to the standard values. Odd-order moments are increased if there is a long tail to the right and decreased if there is a long tail to the left, while even-order moments are increased if either tail is long. A negative value of the coefficient of skewness that the distribution is skewed to the left, or negatively skewed, meaning that the deviations above the mean tend to be smaller than the deviations below the mean, and vice versa. If the coefficent of skewness is close to zero, this could mean symmetry,

Note The fourth moment measures the fatness in the tails, which is always positive. The kurtosis of the standard normal distribution is 3. Using the standard normal distribution as a benchmark, the excess kurtosis of a random variable is defined as the kurtosis minus 3. A higher kurtosis corresponds to a larger extremity of deviations (or outliers), which is called excess kurtosis.

The following diagram compares the shape between the normal distribution and Student's t-distribution. Note that to use the legend with the `stat_function` in ggplot2, we use `scale_colour_manual` along with `colour =` inside the `aes()` as shown below and give names for specific density plots.

```
library(ggplot2)
df <- data.frame(x=seq(-10,10,by=0.1))
ggplot(df) +
    stat_function(aes(x, colour = "dnorm"),fun = dnorm, args = list(mean = 0, sd = 1))  +
    stat_function(aes(x, colour ="dt"),fun = dt, args = list(df = 4)) +
     scale_colour_manual("Legend title", values = c("black", "blue")) +
    labs(x = "x", y = "f(x)",
       title = "Normal Distribution With Mean = 0 & SD = 1") +
  theme(plot.title = element_text(hjust = 0.5))
```



Next we will simulate 10000 samples from a normal distribution with mean 0, and standard deviation 1, then compute and interpret for the skewness and kurtosis, and plot the histogram. Here we also use the function `set.seed()` to set the seed of R's random number generator, this is useful for creating simulations or random objects that can be reproduced.

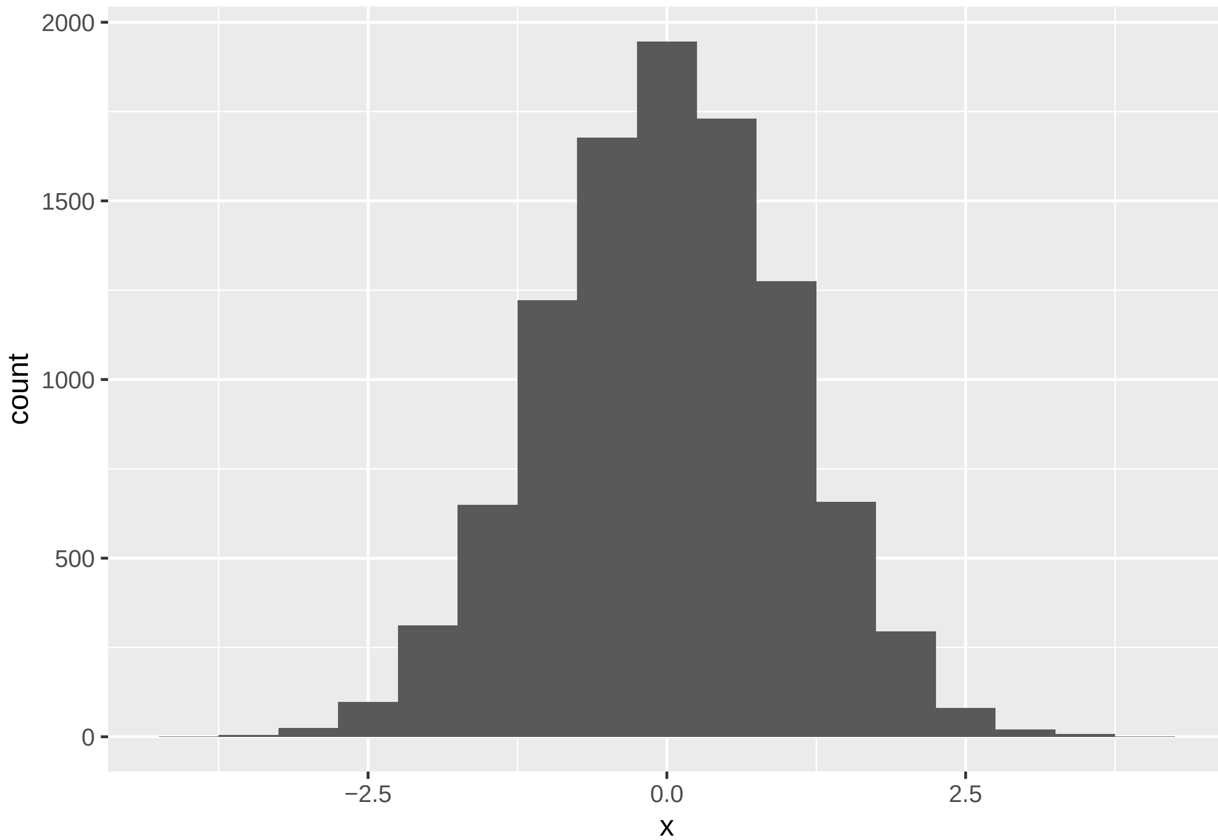
```
set.seed(15) # Set the seed of R's random number generator

#Simulation
n.sample <- rnorm(n = 10000, mean = 0, sd = 1)

#Skewness and Kurtosis
library(moments)
skewness(n.sample)
```

```
## [1] -0.03585812
```

```
kurtosis(n.sample)
```

```
## [1] 2.963189
```

```
ggplot(data.frame(x = n.sample),aes(x)) +
  geom_histogram(binwidth = 0.5)
```




```
set.seed(15)

#Simulation
t.sample <- rt(n = 10000, df = 5)

#Skewness and Kurtosis
library(moments)
skewness(t.sample)
```

```
## [1] 0.06196269
```

```
kurtosis(t.sample)
```

```
## [1] 7.646659
```

```
ggplot(data.frame(x = t.sample),aes(x)) + geom_histogram(binwidth = 0.5)
```