Summary of Chapter 4

1 Concepts

- **Probability density function**: For a continuous random variable X, a probability density function is a function such that

(1) $f(x) \geq 0$

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

(3) $P(a \leq X \leq b) = \int_a^b f(x)dx$ = area under $f(x)$ from a to b for any a, b

- **Cumulative distribution function**: The cumulative distribution function of a continuous random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du \quad \text{for } -\infty < x < \infty$$

- **Summary measures**

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

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

- **Normal Distribution**: The formula for the normal probability density function is

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

  Mean: $E(X) = \mu$

  Variance: $Var(X) = \sigma^2$

- **Standardized Normal Distribution**: The formula for the standardized normal probability density function is

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

  Mean: $\mu = 0$

  Variance: $\sigma^2 = 1$

  If $X \sim N[\mu, \sigma^2]$, $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$, then

$$P(a \le X \le b) = P(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$$

  R function : $P(a \le X \le b) = \text{pnorm}(\frac{b-\mu}{\sigma}) - \text{pnorm}(\frac{a-\mu}{\sigma})$

$$= \text{pnorm}(b, \mu, \sigma) - \text{pnorm}(a, \mu, \sigma).$$

- **Assessing Normality**:

  1. Construct graphs

     (1) stem-and-leaf display

     (2) box-and-whisker plot

     (3) histogram or polygon

  2. Compute descriptive summary measures

     (1) mean, median, and mode (have similar values)

     (2) interquartile range (1.33σ)

     (3) range (6σ)

  3. Use empirical distribution

$$
\begin{aligned}
&P(\mu-\sigma<X<\mu+\sigma)=P(-1<Z<1)=0.68\\
&P(\mu-1.28\sigma<X<\mu+1.28\sigma)=P(-1.28<Z<1.28)=0.80\\
&P(\mu-2\sigma<X<\mu+2\sigma)=P(-2<Z<2)=0.95\\
&P(\mu-3\sigma<X<\mu+3\sigma)=P(-3<Z<3)=0.997
\end{aligned}
$$

  4. Normal probability plot: evaluate the normality according to the linearity of the probability plot.

     R function: qqnorm (D), where D is a dataset.

- **Uniform Distribution**: The formula for the uniform probability density function is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq X \leq b \\ 0 & \text{otherwise} \end{cases}$$

Mean: $\mu = \frac{a+b}{2}$

Variance: $\sigma^2 = \frac{(b-a)^2}{12}$

If $X \sim \text{Uniform}(a, b)$, then

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} \frac{1}{b-a} dx$$

R functions: $P(x_1 \leq X \leq x_2) = \text{punif}(x_2, a, b) - \text{punif}(x_1, a, b).$

- **Exponential distribution**:

The formula for the exponential probability density function is

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

The cumulative distribution of exponential random variable is

$$F(x) = \int_0^x \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}, \qquad F'(x) = f(x)$$

Mean: $\mu = \frac{1}{\lambda}$
Variance: $\sigma^2 = \frac{1}{\lambda^2}$

If $X \sim \exp(\lambda)$, then

$$P(a \leq X \leq b) = \int_a^b \lambda e^{-\lambda u} du = e^{-\lambda a} - e^{-\lambda b}$$

R functions: $P(a \leq X \leq b) = \text{pexp}(b, \lambda) - \text{pexp}(a, \lambda)$.

- **Gamma Distribution**: The formula for the gamma probability density function is

$$f(x) = \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}, \quad x \geq 0$$

with shape parameter $r > 0$ and scale parameter $\lambda > 0$.

Mean: $\mu = \frac{r}{\lambda}$ Variance: $\sigma^2 = \frac{r}{\lambda^2}$

The gamma distribution can assume many different shapes, depending on the values chosen for r and λ.

1. If $r = 1$, the gamma distribution reduces to the exponential distribution with parameter λ.

2. If r is an integer, $x_1, x_2, \cdots, x_r$ are exponential with parameter λ and independent, then $y = x_1 + x_2 + \cdots + x_r$ is distributed as gamma with parameter r and λ.

R functions: If $X \sim \text{Gamma}(r, \lambda)$, then

$$P(a \leq X \leq b) = \text{pgamma}(b, r, \lambda) - \text{pgamma}(a, r, \lambda).$$

- **Weibull Distribution**: The formula for the Weibull probability density function is

$$f(x) = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\theta}\right)^{\beta}\right], \quad x \geq 0$$

where $\beta > 0$ is the shape parameter, and $\theta > 0$ is the scale parameter.

Mean: $\mu = \theta\Gamma\left(1 + \frac{1}{\beta}\right)$

Variance: $\sigma^2 = \theta^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma\left(1 + \frac{1}{\beta}\right)^2\right]$

The Weibull distribution is very flexible, and by appropriate selection of the parameter θ and β, the distribution can assume a wide variety of shapes.

If $\beta = 1$, the Weibull distribution can reduces to the exponential distribution with mean $1/\theta$.

R functions: If $X \sim \text{Weibull}(\theta, \beta)$, then

$$P(a \leq X \leq b) = \text{pweibull}(b, \beta, \theta) - \text{pweibull}(a, \beta, \theta).$$

2 Examples

Example 1. Given a normal distribution with $\mu = 50$ and $\sigma = 4$, what is the probability that

a. $X > 43$?

b. $X < 42$?

c. $42 < X < 48$?

d. $X < 40$ or $X > 55$

a.

$$
\begin{aligned}
P(X > 43) &= P(Z > \frac{43-50}{4}) = P(Z > -1.75) \\
&= 1 - \Phi(-1.75) = 1 - 0.0401 = 0.9599
\end{aligned}
$$

R command: 1-pnorm(43,50,4) or 1-pnorm(-1.75)

b.

$$
\begin{aligned}
P(X < 42) &= P(Z < \frac{42-50}{4}) = P(Z < -2) \\
&= \Phi(-2) = 0.0228
\end{aligned}
$$

R command: pnorm(42,50,4) or 1-pnorm(-2)

c.

$$P(42 < X < 48) == P(\frac{42-50}{4} < Z < \frac{48-50}{4})$$

$$= P(-2 < Z < -0.5) = \Phi(-0.5) - \Phi(-2)$$

$$= 0.3085 - 0.0028 = 0.2857$$

R command: pnorm(48,50,4)-pnorm(42,50,4) or pnorm(-0.5)-pnorm(-2)

d.

$$P(X < 40 \text{ or X} > 55) = P(X < 40) + P(X > 55)$$

$$= P(Z < \frac{40-50}{4}) + P(Z > \frac{55-50}{4})$$

$$= P(Z < -2.5) + P(Z > 1.25)$$

$$= \Phi(-2.5) + 1 - \Phi(1.25)$$

$$= 0.0062 + 1 - 0.8944 = 0.1119$$

R command: pnorm(40,50,4)+1-pnorm(55,50,4) or
pnorm(-2.5)+1-pnorm(1.25)

Example 2. Suppose you sample from a uniform distribution with $a = 0$ and $b = 10$. What is the probability of obtaining a value:

a. between 5 and 7?

b. between 2 and 3?

c. What is the expected value?

d. What is the standard deviation?

a. $P(5 < X < 7) = \frac{7-5}{10-0} = 0.2$

R command: punif(7,0,10)-punif(5,0,10)

b. $P(2 < X < 3) = \frac{3-2}{10-0} = 0.1$

R command: punif(3,0,10)-punif(2,0,10)

c. $\mu = \frac{a+b}{2} = \frac{0+10}{2} = 5$

d. $\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(10-0)^2}{12}} = 2.8868$

Example 3. In a large corporate computer network, user log-ons to the system can be modeled as a Poisson process with a mean of 25 log-ons per hour.

a. What is the probability that there are no log-on in an interval of 6 minutes?

b. What is the probability that the time until the next log-on is between 2 and 3 minutes?

c. Determine the interval of time such that the probability that no log-on occurs in the interval is 0.90.

a. Let X denote the time in hours from the start of the interval until the first log-on. Then, X has an exponential distribution with $\lambda = 25$ log-ons per hour.

$$P(X > 6 \text{ minutes}) = P(X > 0.1 \text{ hours}) = 1 - [1 - e^{-25(0.1)}] = 0.082$$

R command: 1-pexp(0.1,25)

b. $P(0.033 < X < 0.05) = e^{-25(0.033)} - e^{-25(0.05)} = 0.152.$

R command: pexp(0.05,25) - pexp(0.033,25)

c. From $P(X > x) = e^{-25x} = 0.90$, we have -25$x$=ln (0.90)=-0.1054

So, x = 0.00421 hour = 0.25 minute.

Example 4. The time to failure for an electronic subassembly used in RISC workstation is satisfactorily modeled by a Weibull distribution with $\beta = 1/2$ and $\theta = 1000$.

a. What is the mean time to failure?

b. What is the probability of subassemblies expected to survive 4000 hours?

a. The mean time to failure is

$$\begin{aligned} \mu &= \theta\Gamma\left(1+\frac{1}{\beta}\right) = 1000\Gamma\left(1+\frac{1}{1/2}\right) \\ &= 1000\Gamma(3) = 1000 \times 2 = 2000 \text{ hours.} \end{aligned}$$

b. The cumulative Weibull distribution is

$$F(x) = \int_0^x \frac{\beta}{\theta}\left(\frac{u}{\theta}\right)^{\beta-1} \exp\left[-\left(\frac{u}{\theta}\right)^{\beta}\right] du = 1 - \exp\left[-\left(\frac{x}{\theta}\right)^{\beta}\right]$$

$$P(X > 4000) = 1 - P(X < 4000) = \exp\left[-\left(\frac{4000}{1000}\right)^{1/2}\right] = e^{-2} = 0.1353.$$

R command: 1-pweibull(4000,1/2,1000).