CHEM. ENG. II : PROBABILITY AND STATISTICS

Chapter 4. Special Continuous Probability Distributions

4.1 The Exponential Distribution $X \sim Exponential(\lambda)$

Range : $\mathbb{X} = \mathbb{R}^+$
Parameter : $\lambda \in \mathbb{R}^+$
Density function :

$$f_X(x) = \lambda e^{-\lambda x} \qquad x \in \mathbb{R}^+$$

Interpretation : A **continuous** waiting-time model (that is, a useful model for data that arise at the end of a measurement experiment).

The cdf for the exponential distribution can be calculated easily;

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\; dt = \int_0^x \lambda e^{-\lambda t}\; dt = 1 - e^{-\lambda x} \qquad x \geq 0.$$

and note that

$$\mathrm{P}[\ X > x\ ] = 1 - \mathrm{P}[\ X \leq x\ ] = 1 - F_X(x) = e^{-\lambda x}$$

which may give some motivation for using the Exponential model in practice.

Note: Suppose that a mechanical component fails in such a way that the number of failures in a given month has a Poisson distribution with parameter λ, that is, the failure events occur repeatedly through time at random, but failures are relatively rare. Suppose that a component is installed at time $x = 0$. Then the random variable X corresponding to the *time until the first failure* has an exponential distribution, $X \sim Exponential(\lambda)$.

In fact, it can be shown that *all* the inter-failure event times have an exponential distribution, and that the inter-event times are probabilistically independent.

$$\text{EXPECTATION} \qquad \mathrm{E}_{f_X}[\ X\ ] \quad = \int_0^\infty x\lambda e^{-\lambda x}\; dx = \frac{1}{\lambda}$$

$$\text{VARIANCE} \qquad \mathrm{Var}_{f_X}[\ X\ ] \quad = \mathrm{E}_{f_X}[\ X^2\ ] - \{\mathrm{E}_{f_X}[\ X\ ]\}^2 = \int_0^\infty x^2 \lambda e^{-\lambda x}\; dx - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

where the integration by parts is straightforward in both cases.

4.2 The Gamma Distribution $X \sim Gamma(\alpha, \beta)$

Range : $\mathbb{X} = \mathbb{R}^+$
Parameters : $\alpha, \beta \in \mathbb{R}^+$
Density function :

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x \in \mathbb{R}^+$$

where

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\; dt \qquad \alpha > 0.$$

is known as the **Gamma Function**. This integral cannot by calculated analytically, but can be computed using numerical integration for any $\alpha > 0$.

Interpretation : Another **continuous** waiting-time model. It can be shown the sum of i.i.d. Exponential random variables has a Gamma distribution, that is, if $X_1, X_2, ..., X_n$ are independent and identically distributed $Exponential(\lambda)$ random variables, then

$$X = \sum_{i=1}^{n} X_i \sim Gamma(n, \lambda)$$

Notes :

(1) If $\alpha > 1$, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$.

(2) If $\alpha = 1, 2, ...$, $\Gamma(\alpha) = (\alpha-1)!$.

(3) $\Gamma(1/2) = \sqrt{\pi}$.

(4) If $\alpha = 1, 2, ...$, then the $Gamma(\alpha/2, 1/2)$ distribution is known as the **Chi-squared distribution** with α **degrees of freedom**, denoted χ^2_α.

EXAMPLE: Suppose that a mechanical component fails in such a way that the number of failures in a given month has a Poisson distribution with parameter λ. Suppose that a component is installed at time $x = 0$. The random variable X corresponding to the time until the nth ($n = 1, 2, ...$) failure has a gamma distribution, $X \sim Gamma(n, \lambda)$.

The expectation/variance calculations for the Gamma distribution are reasonably straightforward. First consider the expectation of X^r for $r = 1, 2, ...$, that is,

$$\begin{aligned} \mathrm{E}_{f_X}[\, X^r \,] &= \int_{-\infty}^{\infty} x^r f_X(x)\, dx = \int_0^{\infty} x^r \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{r+\alpha-1} e^{-\beta x}\, dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+r)}{\beta^{\alpha+r}} \end{aligned}$$

as the integrand in the last integral is proportional to a Gamma pdf with parameters $r + \alpha$ and β. Hence

$$\mathrm{E}_{f_X}[\, X^r \,] = \frac{\alpha(\alpha+1)...(\alpha+r-1)}{\beta^r}$$

and hence

$$\begin{array}{lll} \text{EXPECTATION} & \mathrm{E}_{f_X}[\, X \,] & = \dfrac{\alpha}{\beta} \\ \text{VARIANCE} & \mathrm{Var}_{f_X}[\, X \,] & = \mathrm{E}_{f_X}[\, X^2 \,] - \{\mathrm{E}_{f_X}[\, X \,]\}^2 = \dfrac{\alpha(\alpha+1)}{\beta^2} - \dfrac{\alpha^2}{\beta^2} = \dfrac{\alpha}{\beta^2} \end{array}$$

4.3 <u>The Beta Distribution</u> $X \sim Beta(\alpha, \beta)$

Range : $\mathbb{X} = (0, 1)$
Parameters : $\alpha, \beta \in \mathbb{R}^+$
Density function :

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad x \in (0, 1).$$

Interpretation : A model for experiments whose outcomes lie in a bounded interval.

EXAMPLE: The percentage content of a particular element in a material is to be measured. The random variable corresponding to the observed content can be modelled using a beta distribution.

The expectation/variance calculations for the Beta distribution are reasonably straightforward, and it can be shown that

$$\begin{array}{lll} \text{EXPECTATION} & \mathrm{E}_{f_X}[\, X \,] & = \dfrac{\alpha}{\alpha+\beta} \\ \text{VARIANCE} & \mathrm{Var}_{f_X}[\, X \,] & = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{array}$$

4.4 <u>The Normal Distribution</u> $X \sim N(\mu, \sigma^2)$

Range : $\mathbb{X} = \mathbb{R}$
Parameters : $\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$
Density function :

$$f_X(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \qquad x \in \mathrm{R}.$$

Interpretation : A probability model that reflects observed (**empirical**) behaviour of data samples; this distribution is often observed in practice.

The pdf is symmetric about μ, and hence μ is controls the *location* of the distribution and σ^2 controls the *spread* or *scale* of the distribution.

Notes :
(1) The Normal density function is justified by the **Central Limit Theorem** (see below).
(2) Special case: $\mu = 0, \sigma^2 = 1$ - the **standard** or **unit** normal distribution. In this case, the density function is denoted $\phi(x)$, and the cdf is denoted $\Phi(x)$ so that

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\ dt = \int_{-\infty}^{x} \left(\frac{1}{2\pi}\right)^{1/2} exp\left\{-\frac{1}{2}t^2\right\}\ dt.$$

This integral can only be calculated numerically (see table).
(3) If $X \sim N(0,1)$, and $Y = \sigma X + \mu$, then $Y \sim N(\mu, \sigma^2)$.
(4) If $X \sim N(0,1)$, and $Y = X^2$, then $Y \sim Gamma(1/2, 1/2) = \chi^2_1$.
(5) If $X \sim N(0,1)$ and $Y \sim \chi^2_\alpha$ are independent random variables, then random variable T, defined by

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

has a **Student-t distribution** with α **degrees of freedom**.
The Student-t distribution plays an important role in certain statistical testing procedures.

4.5 <u>The Central Limit Theorem</u>

The following theorem provides a useful way of approximating probabilities, but also provides a justification for using the Normal distribution in probability and statistics.

<u>THEOREM</u>

Suppose $X_1, ..., X_n$ are i.i.d. random variables with

$$\mathrm{E}_{f_X}[X_i] = \mu, \qquad \mathrm{Var}_{f_X}[X_i] = \sigma^2$$

Let the random variable Z_n be defined by

$$Z_n = \frac{\displaystyle\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}}$$

Then, as $n \longrightarrow \infty$, $Z_n \longrightarrow Z \sim N(0,1)$ **irrespective** of the distribution of $X_1, ..., X_n$, so probability calculations involving Z_n can be approximated using the unit normal c.d.f, Φ, that is,

$$\mathrm{P}[Z_n \leq z] \approx \Phi(z).$$

The central limit theorem is important in probability theory because it justifies the use of an otherwise obscure pdf, and allows us to make approximate probability statements about other probability distributions. For example, can approximate the Binomial, Poisson and Gamma distributions by the normal, that is

$$
\begin{array}{llll}
BINOMIAL & Binomial(n,\theta) & \approx N(n\theta, n\theta(1-\theta)) & n,\theta \text{ large} \\[2ex]
POISSON & Poisson(\lambda) & \approx N(\lambda,\lambda) & \lambda \text{ large} \\[2ex]
GAMMA & Gamma(\alpha,\beta) & \approx N(\alpha/\beta, \alpha/\beta^2) & \alpha \text{ large}
\end{array}
$$

These results allows us to approximate some otherwise awkward probabilities.

EXAMPLE : Suppose that $X \sim Binomial(n,\theta)$. To calculate $\mathrm{P}[X \leq x]$ for large n, first define random variables $X_1, ..., X_n$ to be i.i.d. $Bernoulli(\theta)$. Then, by construction,

$$X = \sum_{i=1}^{n} X_i.$$

and $\mathrm{E}_{f_X}[X_i] = \theta = \mu$, $\mathrm{Var}_{f_X}[X_i] = \theta(1-\theta) = \sigma^2$. Hence, for large n,

$$\mathrm{P}[X \leq x] = \mathrm{P}\left[\sum_{i=1}^{n} X_i \leq x\right] = \mathrm{P}\left[\frac{\sum_{i=1}^{n} X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{x-n\theta}{\sqrt{n\theta(1-\theta)}}\right] \approx \Phi\left(\frac{x-n\theta}{\sqrt{n\theta(1-\theta)}}\right).$$

EXAMPLE : Suppose that $X \sim Poisson(\lambda)$. To calculate $\mathrm{P}[X \leq x]$ for large n, first define random variables $X_1, ..., X_n$ to be i.i.d. $Poisson(\lambda/n)$. Then, by construction,

$$X = \sum_{i=1}^{n} X_i.$$

and $\mathrm{E}_{f_X}[X_i] = \lambda/n = \mu$, $\mathrm{Var}_{f_X}[X_i] = \lambda/n = \sigma^2$. Hence, for large n,

$$\mathrm{P}[X \leq x] = \mathrm{P}\left[\sum_{i=1}^{n} X_i \leq x\right] = \mathrm{P}\left[\frac{\sum_{i=1}^{n} X_i - n\lambda/n}{\sqrt{n\lambda/n}} \leq \frac{x-n\lambda/n}{\sqrt{n\lambda/n}}\right] \approx \Phi\left(\frac{x-\lambda}{\sqrt{\lambda}}\right).$$

EXAMPLE : Suppose that $X \sim Gamma(n,\lambda)$ (for $n = 1, 2, ...$). To calculate $\mathrm{P}[X \leq x]$ for large n, first define random variables $X_1, ..., X_n$ to be i.i.d. $Exponential(\lambda)$. Then, by construction,

$$X = \sum_{i=1}^{n} X_i.$$

and $\mathrm{E}_{f_X}[X_i] = 1/\lambda = \mu$, $\mathrm{Var}_{f_X}[X_i] = 1/\lambda^2 = \sigma^2$. Hence, for large n,

$$\mathrm{P}[X \leq x] = \mathrm{P}\left[\sum_{i=1}^{n} X_i \leq x\right] = \mathrm{P}\left[\frac{\sum_{i=1}^{n} X_i - n/\lambda}{\sqrt{n/\lambda^2}} \leq \frac{x-n/\lambda}{\sqrt{n/\lambda^2}}\right] \approx \Phi\left(\frac{x-n/\lambda}{\sqrt{n/\lambda^2}}\right).$$