Lecture 7: The Normal Probability Distributions

Lecturer: Prof. F. Jalayer, 23 February 2021
Notes prepared by: Prof. F. Jalayer and Dr. H. Ebrahimian

1. Normal Distribution and its statistics

The *Normal* (or *Gaussian* or *Central*) distribution, whose standard PDF is generally denoted as ϕ, is a universally adopted continuous probability distribution. The mean, median and mode of this distribution are identical. The mode of a probability distribution is the point with maximum likelihood (highest probability). The *Standard* Normal PDF for the random variable *U*, has mean 0 and standard deviation equal to 1:

$$f_U\left(u|\mathbf{I}\right)=\phi\left(u|\mathbf{I}\right)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(u)^2} \tag{1}$$

The mean and variance of this PDF can be obtained as:

$$\mathbb{E}\left(u|\mathbf{I}\right)=\int_{-\infty}^{+\infty}u\,\phi\left(u|\mathbf{I}\right)\mathrm{d}u=\int_{-\infty}^{+\infty}\frac{u}{\sqrt{2\pi}}e^{-\frac{1}{2}(u)^2}\mathrm{d}u=\left.\frac{-1}{\sqrt{2\pi}}e^{-\frac{1}{2}(u)^2}\right|_{-\infty}^{+\infty}=0 \tag{2}$$

$$\begin{aligned}\mathbb{E}\left(u^2|\mathbf{I}\right)&=\int_{-\infty}^{+\infty}u^2\,\phi\left(u|\mathbf{I}\right)\mathrm{d}u=\int_{-\infty}^{+\infty}\frac{u^2}{\sqrt{2\pi}}e^{-\frac{1}{2}(u)^2}\mathrm{d}u=\left.\frac{-u}{\sqrt{2\pi}}e^{-\frac{1}{2}(u)^2}\right|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(u)^2}\mathrm{d}u\\&=0+1=1\end{aligned} \tag{3}$$

$$\mathbb{VAR}\left(u|\mathbf{I}\right)=\mathbb{E}\left(u^2|\mathbf{I}\right)-\left[\mathbb{E}\left(u|\mathbf{I}\right)\right]^2=1+0=1 \tag{4}$$

The variable *U* is called the standard normal variable. The *Standard* Normal CDF is defined as:

$$F_U\left(u|\mathbf{I}\right)=P\left(U\le u|\mathbf{I}\right)=\Phi\left(u|\mathbf{I}\right)=\int_{-\infty}^{u}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(v)^2}\mathrm{d}v \tag{5}$$

The above integral cannot be solved analytically. Figure 1 shows the shape of PDF and CDF of the standard Normal distribution.

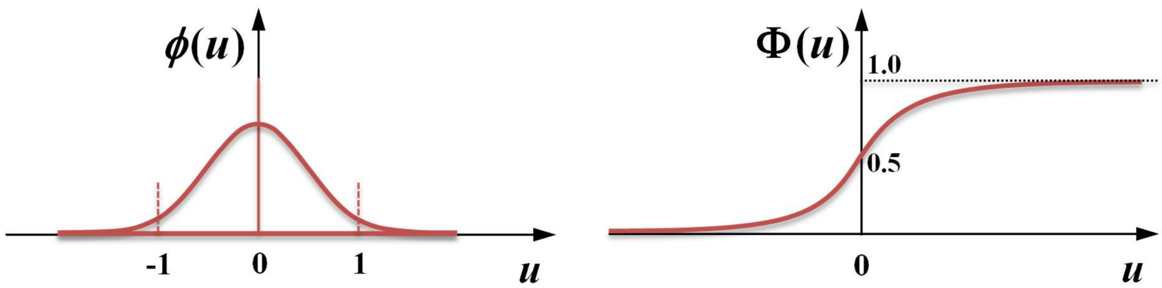

Figure 1: Standard Normal distribution: PDF (left) and CDF (right)

The following properties can be drawn for this distribution:

$$\begin{aligned}\Phi(0) &= P(U \le 0) = 0.50 \\ \Phi(1) &= P(U \le 1) \cong 0.84 \\ \Phi(-1) &= P(U \le -1) \cong 0.16\end{aligned} \tag{6}$$

However, the PDF of Normal distribution can be defined in its general form in terms of the random variable *X*, with the mean μ_X and standard deviation, σ_X, by using the following transformation:

$$U = \frac{X - \mu_X}{\sigma_X} \tag{7}$$

This transformation has specific properties of having zero mean and variance equal to unity for *U*:

$$\begin{aligned}\mathbb{E}(u) &= \mathbb{E}\left(\frac{x-\mu_X}{\sigma_X}\right) = \frac{1}{\sigma_X}\left[\mathbb{E}(x) - \mu_X\right] = 0 \\ \mathbb{VAR}(u) &= \mathbb{VAR}\left(\frac{x-\mu_X}{\sigma_X}\right) = \frac{1}{\sigma_X^2}\mathbb{VAR}(x) = 1\end{aligned} \tag{8}$$

Thus, the PDF can be obtained by:

$$\begin{aligned}f_X(x|\mathbf{I}) &= \frac{\mathrm{d}F_X(x|\mathbf{I})}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}P(X \le x|\mathbf{I}) = \frac{\mathrm{d}}{\mathrm{d}x}P\left(\frac{X-\mu_X}{\sigma_X} \le \frac{x-\mu_X}{\sigma_X}|\mathbf{I}\right) = \frac{\mathrm{d}}{\mathrm{d}x}P(U \le u|\mathbf{I}) \\ &= \frac{\mathrm{d}}{\mathrm{d}x}F_U(u|\mathbf{I}) = \underbrace{\frac{\mathrm{d}F_U(u|\mathbf{I})}{\mathrm{d}u}}_{f_U(u|\mathbf{I})}\frac{\mathrm{d}u}{\mathrm{d}x} = \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(u)^2}\right)\frac{1}{\sigma_X} = \frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}\end{aligned} \tag{9}$$

The Normal CDF can be defined as follows (the integral has cannot be solved analytically):

$$F_X\left(x|\mathbf{I}\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left(\frac{y-\mu_X}{\sigma_X}\right)^2}\mathrm{d}y \tag{10}$$

The uncertain parameter X having a Normal distribution is generally shown as $X \sim \mathbb{N}\left(\mu_X,\sigma_X\right)$.

The PDF of a Normal distribution, and for the sake of comparison, the standard normal PDF are schematically shown in Figure 2.

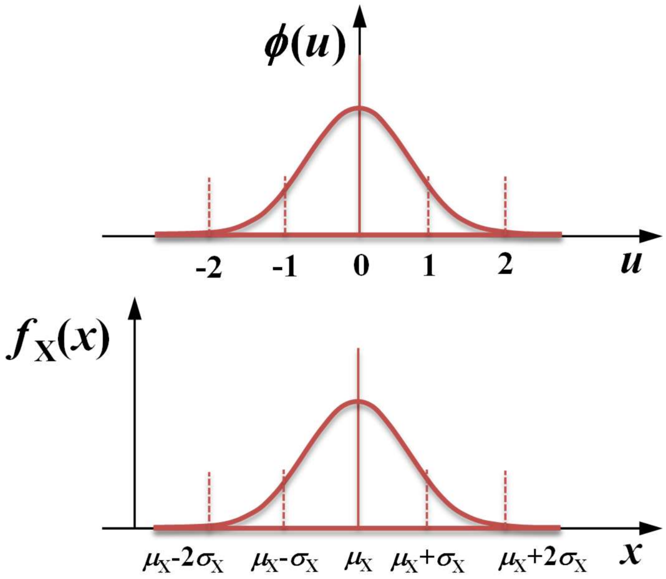

Figure 2: Normal PDF (lower) and standard Normal PDF (upper)

The following properties can be defined:

$$\Phi\left(0\right)=P\left(U\le 0\right)=P\left(\frac{X-\mu_X}{\sigma_X}\le 0\right)=P\left(X\le\mu_X\right)=0.50\Rightarrow p_{50}\equiv\eta_X=\mu_X \tag{11}$$

In addition:

$$\begin{aligned}
&\Phi(1)=P(U\le 1)=P\left(\frac{X-\mu_X}{\sigma_X}\le 1\right)=P\left(X\le \mu_X+\sigma_X\right)\cong 0.84 \Rightarrow p_{84}=\mu_X+\sigma_X \\
&\Phi(-1)=P(U\le -1)=P\left(\frac{X-\mu_X}{\sigma_X}\le -1\right)=P\left(X\le \mu_X-\sigma_X\right)\cong 0.16 \Rightarrow p_{16}=\mu_X-\sigma_X
\end{aligned} \tag{12}$$

Based on the expressions in Eq. (12), one can estimate the σ_X according to Eq. (3). It is also to note that according to Eq. 12,

$$P\left(\mu_X-\sigma_X\le X\le \mu_X+\sigma_X\right)=P\left(x_{16}\le X\le x_{84}\right)\cong 0.68 \tag{13}$$

Hence, with around 70% of probability, the random variable will have a probability content to be within the interval of one standard deviation around the mean. Now, assume that we want to calculate the length of *confidence interval* $\mu_X \pm u_\alpha \sigma_X$ with the probability equal to α , i.e.

$$P\left(\mu_X-u_\alpha\sigma_X\le X\le \mu_X+u_\alpha\sigma_X\right)=\alpha \tag{14}$$

The above expression can be re-written as:

$$\begin{aligned}
&P\left(X\le \mu_X+u_\alpha\sigma_X\right)-P\left(X\le \mu_X-u_\alpha\sigma_X\right)=P\left(U\le u_\alpha\right)-P\left(U\le -u_\alpha\right) \\
&=\Phi\left(u_\alpha\right)-\Phi\left(-u_\alpha\right)=\alpha
\end{aligned} \tag{15}$$

Since the Normal distribution is symmetric around its mean value, it is apparent that:

$$\Phi\left(-u_\alpha\right)=1-\Phi\left(u_\alpha\right) \tag{16}$$

Assuming that $\Phi\left(u_\alpha\right)=p$, Eq. (15) can be written as:

$$p-(1-p)=\alpha \Rightarrow p=\frac{\alpha+1}{2} \tag{17}$$

Thus,

$$u_\alpha=\Phi^{-1}(p)=\Phi^{-1}\left(\frac{\alpha+1}{2}\right) \tag{18}$$

As a result, the confidence interval for having the probability equal to α is:

$$\mu_X - \Phi^{-1}\left(\frac{1+\alpha}{2}\right)\sigma_X \leq X \leq \mu_X + \Phi^{-1}\left(\frac{1+\alpha}{2}\right)\sigma_X \tag{19}$$

2. Lognormal Distribution and its statistics

The random variable X is log-normally distributed if $Y = \ln X$ has normal distribution. Hence, the PDF can be written as:

$$
\begin{aligned}
f_X\left(x\middle|\mathbf{I}\right) &= \frac{\mathrm{d}F_X\left(x\middle|\mathbf{I}\right)}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}P\left(X \le x\middle|\mathbf{I}\right) = \frac{\mathrm{d}}{\mathrm{d}x}P\left(\ln X \le \ln x\middle|\mathbf{I}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}x}P\left(Y \le y\middle|\mathbf{I}\right) = \frac{\mathrm{d}}{\mathrm{d}y}P\left(Y \le y\middle|\mathbf{I}\right)\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}F_Y\left(y\middle|\mathbf{I}\right)}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{x}f_Y\left(y\middle|\mathbf{I}\right) \\
&= \frac{1}{x}\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2} = \frac{1}{\sqrt{2\pi}x\sigma_{\ln X}}e^{-\frac{1}{2}\left(\frac{\ln x-\mu_{\ln X}}{\sigma_{\ln X}}\right)^2}
\end{aligned}
\tag{20}
$$

It is to note that $x > 0$ for the Lognormal distribution; moreover, x in the denominator undermines the symmetry of this distribution in contrast with that of the Normal distribution. In addition, the mean and median are not identical in this distribution. In order to find a relation between the mean $\mu_{\ln X}$ and the median η_X of Log-normally distributed random variable X, i.e. $X \sim \mathbb{LN}\left(\mu_{\ln X}, \sigma_{\ln X}\right)$, one can use the following equalities based on the normal variable Y:

$$
\begin{aligned}
&y_{50} \triangleq \eta_Y \overset{\substack{\text{due to monotonic}\\\text{Mapping}}}{=} \ln\left(x_{50}\right) \triangleq \ln\left(\eta_X\right) \\
&y_{50} \underset{\substack{\text{Normal distribution}\\\text{Property}}}{=} \mu_Y \equiv \mu_{\ln X}
\end{aligned}
\tag{21}
$$

Thus,

$$
\ln\left(\eta_X\right) = \mu_{\ln X} \Rightarrow \eta_X = e^{\mu_{\ln X}} \tag{22}
$$

As a result, the median, mean and mode are not coincide with each other, as shown generally in Figure 3.

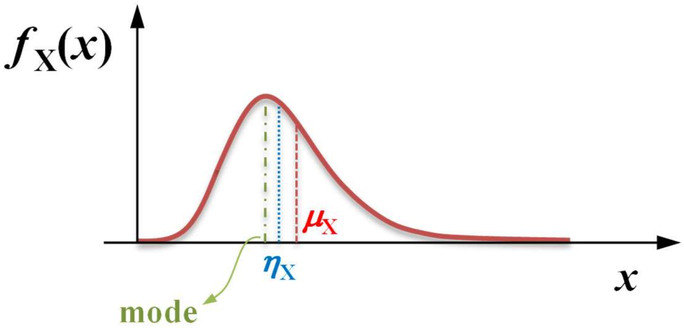

Figure 3: Lognormal PDF

In order to calculate the mean and variance of X, we attempt to estimate the r^{th} moment of X as follows (according to Eq. 20, $f_X\left(x\middle|\mathbf{I}\right)\mathrm{d}x = f_Y\left(x\middle|\mathbf{I}\right)\mathrm{d}y$):

$$\mathbb{E}\left[X^r\left|\mathbf{I}\right.\right] = \mathbb{E}\left[e^{r\ln X}\left|\mathbf{I}\right.\right] = \int_{-\infty}^{\infty} e^{r\ln x} f_X\left(x\middle|\mathbf{I}\right)\mathrm{d}x = \int_{-\infty}^{\infty} e^{ry} f_Y\left(x\middle|\mathbf{I}\right)\mathrm{d}y \tag{23}$$

Hence,

$$\begin{aligned}
\mathbb{E}\left[X^r\left|\mathbf{I}\right.\right] &= \int_{-\infty}^{\infty} e^{ry}\frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}\mathrm{d}y = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma_Y} e^{-\frac{1}{2}(-2ry)} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma_Y}\exp\left[-\frac{1}{2}\left(\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\,ry\right)\right]\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\,\sigma_Y}\exp\left[-\frac{1}{2}\left(\frac{\left(y-\mu_Y\right)^2 - 2y\cdot r\sigma_Y^2 \pm \left(r\sigma_Y^2\right)^2 \pm 2\mu_Y\cdot r\sigma_Y^2}{\sigma_Y^2}\right)\right]\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\exp\left(\frac{1}{2}r^2\sigma_Y^2\right)\cdot\exp\left(r\mu_Y\right)\cdot\frac{1}{\sqrt{2\pi}\,\sigma_Y}\exp\left[-\frac{1}{2}\left(\frac{y-\left[\mu_Y + r\sigma_Y^2\right]}{\sigma_Y}\right)^2\right]\mathrm{d}y \\
&= \exp\left(\frac{1}{2}r^2\sigma_Y^2\right)\cdot\exp\left(r\mu_Y\right) = \exp\left(\frac{1}{2}r^2\sigma_{\ln X}^2\right)\cdot\exp\left(r\ln\eta_X\right) = \left(\eta_X\right)^r e^{\frac{1}{2}r^2\sigma_{\ln X}^2}
\end{aligned} \tag{24}$$

As a result, the expected value of X can be defined as:

$$\mathbb{E}\left[X\left|\mathbf{I}\right.\right] \equiv \mu_X = \eta_X e^{\frac{1}{2}\sigma_{\ln X}^2} \Rightarrow \mu_X = e^{\mu_{\ln X} + \frac{1}{2}\sigma_{\ln X}^2} \tag{25}$$

The variance can be estimated as:

$$\begin{aligned}\mathbb{VAR}(X) &= \mathbb{E}\left[X^2 \middle| \mathbf{I}\right] - \left(\mathbb{E}\left[X \middle| \mathbf{I}\right]\right)^2 = e^{2\mu_{\ln X}+2\sigma_{\ln X}^2} - e^{2\mu_{\ln X}+\sigma_{\ln X}^2} = e^{2\mu_{\ln X}+\sigma_{\ln X}^2}\left(e^{\sigma_{\ln X}^2}-1\right) \\ &= \left(\mu_X\right)^2\left(e^{\sigma_{\ln X}^2}-1\right)\end{aligned} \tag{26}$$

Based on Eq. (26), the *coefficient of variation* of *X*, denoted as V_X, can be estimated as:

$$V_X = \frac{\sigma_X}{\mu_X} = \sqrt{e^{\sigma_{\ln X}^2}-1} \tag{27}$$

In case that $\sigma_{\ln X}$ is small,

$$V_X = \sqrt{e^{\sigma_{\ln X}^2}-1} = \sqrt{1+\sigma_{\ln X}^2+\frac{\left(\sigma_{\ln X}^2\right)^2}{2!}+\cdots-1} \cong \sqrt{1+\sigma_{\ln X}^2-1} \cong \sigma_{\ln X} \tag{28}$$

It is noteworthy that $\sigma_{\ln X}$ has no units, and X can be represented as $X \sim \mathbb{LN}\left(\eta_X, V_X\right)$. Equivalently, with reference to Eq. (25) and Eq. (27), the parameters of lognormal distribution, i.e. $\mu_{\ln X}$ and $\sigma_{\ln X}$ can be obtained if the expected value and variance of *X* are known:

$$\begin{aligned}\sigma_{\ln X} &= \sqrt{\ln\left(1+\frac{\sigma_X^2}{\mu_X^2}\right)} \\ \mu_{\ln X} &= \ln \mu_X - \frac{1}{2}\sigma_{\ln X}^2\end{aligned} \tag{29}$$

In addition, the parameters of the lognormal distribution (i.e. the mean and standard deviation of ln*X*) can also be estimated according to the ordered statistics as follows:

$$\begin{aligned}\mu_{\ln X} &= \ln\left(\eta_X\right) = \ln\left(x_{50}\right) \\ \sigma_{\ln X} &= \frac{\ln x_{84} - \ln x_{16}}{2} = \frac{1}{2}\ln\left(\frac{x_{84}}{x_{16}}\right) \equiv \frac{\ln\left(x_{84}/x_{50}\right)+\ln\left(x_{50}/x_{16}\right)}{2}\end{aligned} \tag{30}$$