$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right],$$

for all $(x_1, x_2) \in \mathbb{R}^2$. To check that f_{X_1,X_2}, as above, is a p.d.f., first note that $f_{X_1,X_2}(x_1, x_2) \geq 0, \forall (x_1, x_2) \in \mathbb{R}^2$. Now, changing variables to $y_1 = \frac{x_1-\mu_1}{\sigma_1}, y_2 = \frac{x_2-\mu_2}{\sigma_2}$, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2 \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{y_2^2}{2}\right)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}(y_1-\rho y_2)^2\right] dy_1 dy_2 \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{y_2^2}{2}\right) dy_2 \\
&= 1.
\end{aligned}$$

This completes the verification that f_{X_1,X_2} is a joint p.d.f..

Remark 1.384. Let $X = (X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ for some $\mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}, \sigma_1 > 0, \sigma_2 > 0, \rho \in (-1, 1)$.

(a) The marginal p.d.f. of X_2 is given by

$$\begin{aligned}
&f_{X_2}(x_2) \\
&= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_1, \forall x_2 \in \mathbb{R} \\
&= \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] \\
&\times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sigma_1^2(1-\rho^2)}\left\{x_1 - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right)\right\}^2\right] dx_1 \\
&= \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]
\end{aligned}$$

and hence $X_2 \sim N(\mu_2, \sigma_2^2)$. Similarly, $X_1 \sim N(\mu_1, \sigma_1^2)$. Thus the parameters $\mu_1 = \mathbb{E}X_1, \sigma_1^2 = Var(X_1), \mu_2 = \mathbb{E}X_2, \sigma_2^2 = Var(X_2)$ have their own interpretation.

(b) The covariance $Cov(X_1, X_2)$ is given by

$$\begin{aligned} Cov(X_1, X_2) &= \mathbb{E}(X_1 - \mu_1)(X_2 - \mu_2) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2) f_{X_1,X_2}(x_1, x_2)\, dx_1 dx_2 \end{aligned}$$

By changing variables to $y_1 = \frac{x_1-\mu_1}{\sigma_1}, y_2 = \frac{x_2-\mu_2}{\sigma_2}$, and simplifying the above expression, we have $Cov(X_1, X_2) = \rho\sigma_1\sigma_2$. Consequently, the correlation $\rho(X_1, X_2) = \rho$. We now have the interpretation of the parameter ρ.

(c) The conditional distribution of X_1 given $X_2 = x_2 \in \mathbb{R}$ is described by the conditional p.d.f.

$$\begin{aligned} f_{X_1|X_2}(x_1 \mid x_2) &= \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sigma_1^2(1-\rho^2)}\left\{x_1 - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)\right)\right\}^2\right], \forall x_1 \in \mathbb{R} \end{aligned}$$

and hence $X_1 \mid X_2 = x_2 \sim N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1 - \rho^2))$. Similarly, for $x_1 \in \mathbb{R}$, $X_2 \mid X_1 = x_1 \sim N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1 - \rho^2))$.

(d) Using the conditional distributions obtained above, we conclude

$$\begin{aligned} \mathbb{E}[X_1 \mid X_2 = x_2] &= \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \\ Var[X_1 \mid X_2 = x_2] &= \sigma_1^2(1 - \rho^2), \\ \mathbb{E}[X_2 \mid X_1 = x_1] &= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \\ Var[X_2 \mid X_1 = x_1] &= \sigma_2^2(1 - \rho^2). \end{aligned}$$

(e) If X_1 and X_2 are independent, then they are uncorrelated and in particular $\rho = \rho(X_1, X_2) = 0$. Conversely, if $\rho = 0$, then

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2}\left\{\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right\}\right] \\ &= f_{X_1}(x_1) f_{X_2}(x_2), \forall (x_1, x_2) \in \mathbb{R}^2 \end{aligned}$$

and hence X_1 and X_2 are independent.

(f) Consider the matrix

$$\Sigma = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & Var(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

This is usually called the Covariance matrix of $X = (X_1, X_2)$. Alternative terminology such as Variance-Covariance matrix or Dispersion matrix is also used. Observe that this is a symmetric matrix with $det(\Sigma) = \sigma_1^2\sigma_2^2(1-\rho^2) > 0$ and hence the matrix is invertible. In fact, this matrix is positive-definite and its eigen-values are positive. The joint p.d.f. can be rewritten as

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left[-\frac{1}{2}(x_1 - \mu_1, x_2 - \mu_2)\Sigma^{-1}\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right], \forall (x_1, x_2) \in \mathbb{R}^2.$$

(g) We now compute the joint MGF of X. We have

$$\begin{aligned}
&M_X(t_1, t_2) \\
&= \mathbb{E}\exp(t_1X_1 + t_2X_2) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(t_1x_1 + t_2x_2) f_{X_1,X_2}(x_1, x_2)\, dx_1dx_2 \\
&= \int_{-\infty}^{\infty} \exp(t_2x_2) f_{X_2}(x_2) \int_{-\infty}^{\infty} \exp(t_1x_1) f_{X_1|X_2}(x_1 \mid x_2)\, dx_1dx_2 \\
&= \int_{-\infty}^{\infty} \exp(t_2x_2) f_{X_2}(x_2) \exp(\mu_1t_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)t_1 + \frac{1}{2}\sigma_1^2(1-\rho^2)t_1^2)\, dx_2 \\
&= \exp(\mu_1t_1 + \rho\frac{\sigma_1}{\sigma_2}(-\mu_2)t_1 + \frac{1}{2}\sigma_1^2(1-\rho^2)t_1^2) \int_{-\infty}^{\infty} \exp(t_2x_2 + \rho\frac{\sigma_1}{\sigma_2}t_1x_2) f_{X_2}(x_2)\, dx_2 \\
&= \exp(\mu_1t_1 + \rho\frac{\sigma_1}{\sigma_2}(-\mu_2)t_1 + \frac{1}{2}\sigma_1^2(1-\rho^2)t_1^2) \exp\left(\mu_2\left\{t_2 + \rho\frac{\sigma_1}{\sigma_2}t_1\right\} + \frac{1}{2}\sigma_2^2\left\{t_2 + \rho\frac{\sigma_1}{\sigma_2}t_1\right\}^2\right) \\
&= \exp(\mu_1t_1 + \mu_2t_2 + \frac{1}{2}\sigma_1^2t_1^2 + \frac{1}{2}\sigma_2^2t_2^2 + \rho\sigma_1\sigma_2t_1t_2), \forall (t_1, t_2) \in \mathbb{R}^2.
\end{aligned}$$

(h) Let $c_1, c_2 \in \mathbb{R}$ such that at least one of c_1, c_2 is not zero and take $Y = c_1X_1 + c_2X_2$. Now,

$$M_Y(t) = \mathbb{E}\exp(c_1tX_1 + c_2tX_2) = \exp\left[(\mu_1c_1 + \mu_2c_2)t + \left(\frac{1}{2}\sigma_1^2c_1^2 + \frac{1}{2}\sigma_2^2c_2^2 + \rho\sigma_1\sigma_2c_1c_2\right)t^2\right], \forall t \in \mathbb{R}.$$

Looking at the structure of the MGF, we conclude that $Y = c_1X_1 + c_2X_2 \sim N(c_1\mu_1 + c_2\mu_2, c_1^2\sigma_1^2 + c_2\sigma_2^2 + 2\rho c_1c_2\sigma_1\sigma_2)$.

Remark 1.385. The above statement for the linear combination of X_1, X_2 actually characterizes the bivariate Normal distribution. If $X = (X_1, X_2)$ is such that $\mathbb{E}X_1 = \mu_1, \mathbb{E}X_2 = \mu_2, Var(X_1) = \sigma_1^2 > 0, Var(X_2) = \sigma_2^2 > 0, \rho(X_1, X_2) = \rho \in (-1, 1)$ and $c_1X_1 + c_2X_2 \sim N(c_1\mu_1 + c_2\mu_2, c_1^2\sigma_1^2 + c_2\sigma_2^2 + 2\rho c_1c_2\sigma_1\sigma_2)$ for all $(c_1, c_2) \neq (0, 0)$, then

$$\begin{aligned} M_X(t_1, t_2) &= \mathbb{E}\exp(t_1X_1 + t_2X_2) \\ &= M_{t_1X_1+t_2X_2}(1) \\ &= \exp(\mu_1t_1 + \mu_2t_2 + \frac{1}{2}\sigma_1^2t_1^2 + \frac{1}{2}\sigma_2^2t_2^2 + \rho\sigma_1\sigma_2t_1t_2), \forall(t_1, t_2) \in \mathbb{R}^2. \end{aligned}$$

Since an MGF determines the distribution, we conclude $X = (X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

Remark 1.386 (Interpretation of parameters appearing in the p.d.f. of a Continuous RV). In the examples of continuous RVs discussed in this course, we have seen that certain parameters appear in the description of p.d.fs. If we specify the values of these parameters, then we obtain a specific example of distribution from a family of possible distributions. In certain cases, we have already been able to interpret them in terms of properties of the distribution of the RV. For example, if $X \sim N(\mu, \sigma^2)$, then $\mu = \mathbb{E}X$ and $\sigma^2 = Var(X)$. We list some interpretation of these parameters.

(a) (Location parameter) If we have a family of p.d.f.s $f_\theta, \theta \in \Theta$, where θ is a real valued parameter (i.e. $\Theta \subseteq \mathbb{R}$) and if $f_\theta(x) = f_0(x - \theta), \forall x \in \mathbb{R}$, then we say that θ is a location parameter for the family of distributions given by the p.d.f.s f_θ. In this case, the family is called a location family and the p.d.f. f_0 is free of θ, i.e. does not depend on θ. We can restate this fact in terms of the corresponding RVs X_θ as follows: the p.d.f./distribution of $X_\theta - \theta$ does not depend on θ.

(b) (Scale parameter) If we have a family of p.d.f.s f_θ, where θ is a real valued parameter (i.e. $\Theta \subseteq \mathbb{R}$) and if $f_\theta(x) = \frac{1}{\theta}f_1(\frac{x}{\theta}), \forall x \in \mathbb{R}$, then we say that θ is a scale parameter for the family of distributions given by the p.d.f.s f_θ. In this case, the family is called a scale family and the p.d.f. f_1 is free of θ, i.e. does not depend on θ. We can restate this fact in terms of the corresponding RVs X_θ as follows: the p.d.f./distribution of $\frac{1}{\theta}X_\theta$ does not depend on θ.

(c) (Location-scale parameter) If we have a family of p.d.f.s $f_{\mu,\sigma}$ with $\sigma > 0$ and if $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right) = f_{0,1}(x), \forall x \in \mathbb{R}$, then we say that (μ, σ) is a location-scale parameter for the family of distributions given by the p.d.f.s $f_{\mu,\sigma}$. In this case, the family is called a location-scale family and the p.d.f. $f_{0,1}$ is free of (μ, σ), i.e. does not depend on (μ, σ). We can restate this fact in terms of the corresponding RVs $X_{\mu,\sigma}$ as follows: the p.d.f./distribution of $\frac{X_{\mu,\sigma}-\mu}{\sigma}$ does not depend on (μ, σ).

(d) (Shape parameter) Some family of p.d.f.s also has a shape parameter, where changing the value of the parameter affects the shape of the graph of the p.d.f..

Example 1.387. (a) The family of RVs $X_{\mu,\theta} \sim \mathit{Cauchy}(\mu, \theta), \mu \in \mathbb{R}, \theta > 0$ with the p.d.f.

$$f_{\mu,\theta}(x) = \frac{\theta}{\pi} \frac{1}{\theta^2 + (x-\mu)^2}, \forall x \in \mathbb{R}$$

is a location-scale family with location parameter μ and scale parameter θ.

(b) For the family of RVs $X_\alpha \sim \mathit{Gamma}(\alpha, 1), \alpha > 0$ with the p.d.f.

$$f_\alpha(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}, \text{ if } x > 0, \\ 0, \text{ otherwise} \end{cases}$$

α is a shape parameter.

Definition 1.388 (Weibull distribution)**.** We say that an RV X follows the Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, if its p.d.f. is given by

$$f_X(x) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right], \text{ if } x > 0, \\ 0, \text{ otherwise.} \end{cases}$$

Note 1.389. Let $X \sim \mathit{Exponential}(\beta^\alpha)$ for some $\alpha, \beta > 0$. Then $Y = X^{\frac{1}{\alpha}}$ follows the Weibull distribution with shape parameter α and scale parameter β.

Definition 1.390 (Pareto distribution)**.** We say that an RV X follows the Pareto distribution with scale parameter $\theta > 0$ and shape parameter $\alpha > 0$, if its p.d.f. is given by

$$f_X(x) = \begin{cases} \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Remark 1.391 (Descriptive Measures of Probability Distributions)*.* The distribution of an RV provides numerical values through which we can quantify/understand the manner in which the RV takes values in various subsets of the real line. However, at times, it is difficult to grasp the features of the RV from the distribution. As an alternative, we typically use four types of numerical quantities associated with the distribution to summarize the information. We refer to them as descriptive measures of the probability distribution.

(a) Measures of Central Tendency or location: here, we try to find a 'central' value around which the possible values of the RV are distributed.

(b) Measures of Dispersion: once we have an idea of the 'central' value of the RV (equivalently, the probability distribution), we check the scattering/dispersion of the all the possible values of the RV around this 'central' value.

(c) Measures of Skewness: here, we try to quantify the asymmetry of the probability distribution.

(d) Measures of Kurtosis: here, we try to measure the thickness of the tails of the RV (equivalently, the probability distribution) while comparing with the Normal distribution.

We describe these measures along with examples.

Example 1.392 (Measures of Central Tendency)**.** (a) The Mean of an RV is a good example of a measure of central tendency. It also has the useful property of linearity. However, it may be affected by few extreme values, referred to as the outliers. The mean may not exist for all distributions.

(b) Median, i.e. a quantile of order $\frac{1}{2}$ of an RV is always defined and is usually not affected by a few outliers. However, the median lacks the linearity property, i.e. a median of $X + Y$ has no general relationship with the medians of X and Y. Further, a median focuses on the probabilities with which the values of the RV occur rather than the exact numerical values. A median need not be unique.

(c) The mode m_0 of a probability distribution is the value that occurs with 'highest probability', and is defined by $f_X(m_0) = \sup\{f_X(x) : x \in S_X\}$, where f_X denotes the p.m.f./p.d.f. of X, as appropriate and S_X denotes the support of X. Mode need not be unique. Distributions with one, two or multiple modes are called unimodal, bimodal or multimodal distributions, respectively. Usually, it is easy calculate. However, it may so happen that a distribution has more than multiple modes situated far apart, in which case it may not be suitable for a measure of central tendency.

Example 1.393 (Measures of Dispersion)**.** (a) If the support S_X of an RV X is contained in the interval $[a, b]$ and this is the smallest such interval, then we define $b-a$ to be the range of X. This measure of dispersion does not take into account the probabilities with which the values of X are distributed.

(b) Mean Deviation about a point $c \in \mathbb{R}$: If $\mathbb{E}|X - c|$ exists, we define it to be the mean deviation of X about the point c. Usually, we take c to be the mean (if it exists) or the median and obtain mean deviation about the mean or median, respectively. However, it may be difficult to compute and even may not exist. The mean deviations are also affected by a few outliers.

(c) Standard Deviation: As defined earlier, the standard deviation of an RV X is $\sqrt{Var(X)}$, if it exists. Compared to the mean deviation, the standard deviation is usually easier to compute. The standard deviation is affected by a few outliers.

(d) Quartile Deviation: Recall that $\mathfrak{z}_{0.25}$ and $\mathfrak{z}_{0.75}$ denotes the lower and upper quartiles. We define $\mathfrak{z}_{0.75} - \mathfrak{z}_{0.25}$ to be the inter-quartile range and refer to $\frac{1}{2}[\mathfrak{z}_{0.75} - \mathfrak{z}_{0.25}]$ as the semi-inter-quartile range or the quartile deviation. This measures the spread in the middle half of the distribution and is therefore not influenced by extreme values. However, it does not take into account the numerical values of the RV.

(e) Coefficient of Variation: The coefficient of variation of X is defined as $\frac{\sqrt{Var(X)}}{\mathbb{E}X}$, provided $\mathbb{E}X \neq 0$. This aims to measure the variation per unit of mean. It, by definition, does not depend on the unit of measurement. However, it may be sensitive to small changes in the mean, if it is close to zero.

Note 1.394 (A Measure of Skewness)**.** If the distribution of an RV X is symmetric about the mean μ, then $f_X(\mu + x) = f_X(\mu - x), \forall x \in \mathbb{R}$, where f_X denotes the p.m.f./p.d.f. of X. If this is not the case, then two cases may occur.

(a) (Positively skewed) the distribution may have more probability mass towards the right hand side of the graph of f_X. In this case, the tails on the right hand side are longer.
(b) (Negatively skewed) the distribution may have more probability mass towards the left hand side of the graph of f_X. In this case, the tails on the left hand side are longer.

To measure this asymmetry, we usually look at $\mathbb{E}Z^3$, where $Z = \frac{X-\mathbb{E}X}{\sqrt{Var(X)}}$, provided the moments exist. Note that Z is independent of the units of measurement and

$$\mathbb{E}Z^3 = \frac{\mathbb{E}(X - \mathbb{E}X)^3}{(Var(X))^{\frac{3}{2}}} = \frac{\mu_3(X)}{(\mu_2(X))^{3/2}}.$$

We may refer to a distribution being positively or negatively skewed according as the above quantity being positive or negative. If $X \sim \mathit{Exponential}(\lambda)$, then $\mathbb{E}Z^3 = 2$ and hence the distribution of X is positively skewed.

Note 1.395. There are many other measures of skewness used in practice. However, we do not discuss them in this course.

Note 1.396 (A measure of Kurtosis)**.** The probability distribution of X is said to have higher (respectively, lower) kurtosis than the Normal distribution, if its p.m.f./p.d.f., in comparison with the p.d.f. of a Normal distribution, has a sharper (respectively, rounded) peak and longer/fatter (respectively, shorter/thinner) tails. To measure the kurtosis of X, we look at $\mathbb{E}Z^4$, where $Z = \frac{X-\mathbb{E}X}{\sqrt{Var(X)}}$, provided the moments exist. Note that Z is independent of the units of measurement and

$$\mathbb{E}Z^4 = \frac{\mathbb{E}(X - \mathbb{E}X)^4}{(Var(X))^2} = \frac{\mu_4(X)}{(\mu_2(X))^2}.$$

If $X \sim N(\mu, \sigma^2)$, then $Z \sim N(0, 1)$ and hence $\mathbb{E}Z^4 = 3$ (see Remark 1.248). For a general RV X, the quantity $\frac{\mu_4(X)}{(\mu_2(X))^2} - 3$ is referred to as the excess kurtosis of X. If the excess kurtosis is zero, positive or negative, then we refer to the corresponding probability distribution as mesokurtic, leptokurtic or platykurtic, respectively. If $X \sim \mathit{Exponential}(\lambda)$, then $\mathbb{E}Z^4 = 9$ and hence the distribution of X is leptokurtic.

Definition 1.397 (Quantile function of an RV)**.** Let X be an RV with the DF F_X. The function $Q_X : (0,1) \to \mathbb{R}$ defined by

$$Q_X(p) := \inf\{x \in \mathbb{R} : F_X(x) \geq p\}, \forall p \in (0,1)$$

is called the quantile function of X.

Proposition 1.398 (Probability integral transform)**.** *Let X be a continuous RV with the DF F_X, p.d.f. f_X and quantile function Q_X.*

(a) We have $F_X(X) \sim Uniform(0,1)$.
(b) For any $U \sim Uniform(0,1)$, we have $Q_X(U) \stackrel{d}{=} X$.

Proof. We prove only the first statement. The proof of the second statement is similar. Take $Y = F_X(X)$. Then,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(F_X(X) \leq y) = \begin{cases} 0, & \text{if } y < 0, \\ 1, & \text{if } y \geq 1. \end{cases}$$

For $y \in [0,1)$, we have

$$\mathbb{P}(F_X(X) = y) = \mathbb{P}(x_1 \leq X \leq x_2) = 0$$

for some $x_1, x_2 \in \mathbb{R}$ with $F_X(x_1) = F_X(x_2)$. Here, we have used the fact that X is a continuous RV. Now, for $y \in [0,1)$,

$$\begin{aligned}
\mathbb{P}(F_X(X) \leq y) &= \mathbb{P}(F_X(X) < y) \\
&= 1 - \mathbb{P}(F_X(X) \geq y) \\
&= 1 - \mathbb{P}(X \geq Q_X(y)) \\
&= 1 - \mathbb{P}(X > Q_X(y))
\end{aligned}$$

$$= \mathbb{P}(X \leq Q_X(y))$$
$$= F_X(Q_X(y)$$
$$= y.$$

Hence, $Y = F_X(X) \sim Uniform(0,1)$. This completes the proof. $\square$

Note 1.399. Let X be an RV with the quantile function Q_X. If we can generate random samples $U_1, U_2, \cdots, U_n$ from $U \sim Uniform(0,1)$, then $Q_X(U_1), Q_X(U_2), \cdots, Q_X(U_n)$ are random samples from the distribution of X. This observation may be used in practice to generate random samples for known distributions from the $Uniform(0,1)$ distribution.

Note 1.400 (Moments do not determine the distribution of an RV)**.** Let $X \sim N(0,1)$ and consider $Y = e^X$. The distribution of Y is usually called the lognormal distribution, since $\ln Y = X \sim N(0,1)$. Using standard techniques, we can compute the p.d.f. of Y:

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-1} \exp\left[-\frac{(\ln y)^2}{2}\right], \text{ if } y > 0, \\ 0, \text{ otherwise.} \end{cases}$$

It can be shown that the continuous RVs $X_\alpha, \alpha \in [-1,1]$ with the p.d.fs

$$f_{X_\alpha}(y) = f_Y(y)\left[1 + \alpha \sin(2\pi \ln y)\right], \forall y \in \mathbb{R}$$

has the same moments as Y. However, the distributions are different. This shows that the moments of an RV do not determine the distribution. (see the article 'On a property of the lognormal distribution' by C.C. Heyde, published in Journal of the Royal Statistical Society: Series B, volume 29 (1963).)

Note 1.401 (Operations on DFs)**.** Recall that a DF $F : \mathbb{R} \to [0,1]$ is characterized by the properties that it is right continuous, non-decreasing and $\lim_{x\to\infty} F(x) = 1, \lim_{x\to-\infty} F(x) = 0$. Given two DFs $F, G : \mathbb{R} \to [0,1]$ and $\alpha \in [0,1]$, we make the following observations.

(a) (Convex combination of DFs) The function $H : \mathbb{R} \to [0,1]$ defined by $H(x) := \alpha F(x) + (1-\alpha)G(x), \forall x \in \mathbb{R}$ has the relevant properties and hence is a DF.