# Chapter 4

# RANDOM VARIABLES

## 4.1 INTRODUCTION

**What Is Random Variable?**
We are dealing with

- a physical phenomenon which exhibits randomness.
- the outcome of any one occurence (trial) cannot be predicted.
- the probability of any subset of possible outcomes is well-defined.

We ascribe the term *random variable* to such a phenomenon. Note that a random variable is not defined by a specific number; rather it is defined by the probabilities of all subsets of the possible outcomes. An outcome of a particular trial is called a *realization* of the random variable.

An example is outcome of rolling a dice. Let $x$ represent the outcome (not of a particular trial, but in general). Then, $x$ is not represented by a single outcome, but is defined by the set of possible outcomes ($\{1, 2, 3, 4, 5, 6\}$) and the probability of the possible outcome(s) (1/6 each). When we say $x$ is 1 or 2 or so on, we really should say a realization of $x$ is such.

A random variable can be discrete or continuous. If the outcome of a random variable belongs to a discrete space, the random variable is *discrete*. An example is the outcome of rolling a dice. On the other hand, if the outcome belongs to a continuous space, the random variable is *continuous*. For instance, composition or temperature of a distillation column can be viewed as continuous random variables.

**What Is Statistics?**

*Statistics* deals with the application of probability theory to real problems. There are two basic problems in statistics.

- Given a probabilistic model, predict the outcome of future trial(s). For instance one may say:

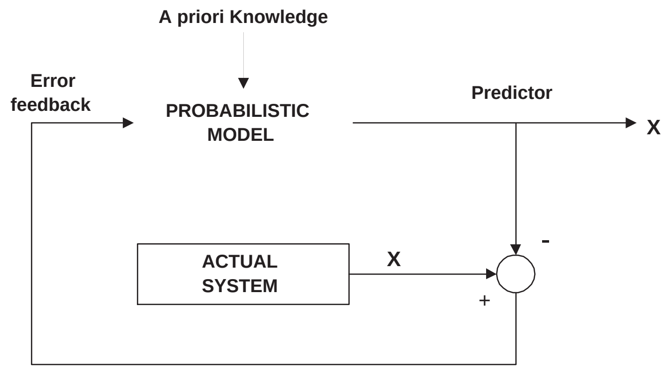
  > choose the prediction $\hat{x}$ such that expected value of $(x - \hat{x})^2$ is minimized.

- Given collected data, define / improve a probabilistic model.

  For instance, there may be some unknown parameters (say $\theta$) in the probabilistic model. Then, given data $X$ generated from the particular probabilistic model, one should construct an estimate of $\theta$ in the form of $\hat{\theta}(X)$. For example, $\hat{\theta}(X)$ may be constructed based on the objective of minimizing expected value of $\|\theta - \hat{\theta}\|_2^2$.

  Another related topic is *hypothesis testing*, which has to do with testing whether a given hypothesis is correct (i.e, how correct defined in terms of probability), based on available data.

In fact, one does both. That is, as data come in, one may continue to improve the probabilistic model and use the updated model for further prediction.

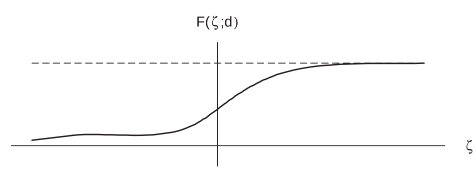



## 4.2 BASIC PROBABILITY CONCEPTS

### 4.2.1 PROBABILITY DISTRIBUTION, DENSITY: SCALAR CASE

A random variable is defined by a function describing the probability of the outcome rather than a specific value. Let $d$ be a *continuous* random variable ($d \in \mathcal{R}$). Then one of the following functions is used to define $d$:

- **Probability Distribution Function**
  The probability distribution function $F(\zeta; d)$ for random variable $d$ is defined as
  $$F(\zeta; d) = \Pr\{d \leq \zeta\} \tag{4.1}$$

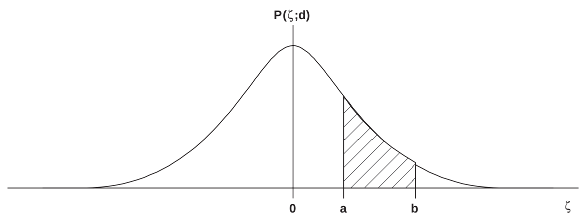



  where Pr denotes the probability. Note that $F(\zeta; d)$ is monotonically increasing with $\zeta$ and asymptotically reaches 1 as $\zeta$ approaches its upper limit.

- **Probability Density Function**

  The probability density function $\mathcal{P}(\zeta; d)$ for random variable $d$ is defined as

$$\mathcal{P}(\zeta; d) = \frac{dF(\zeta; d)}{d\zeta} \tag{4.2}$$




Note that

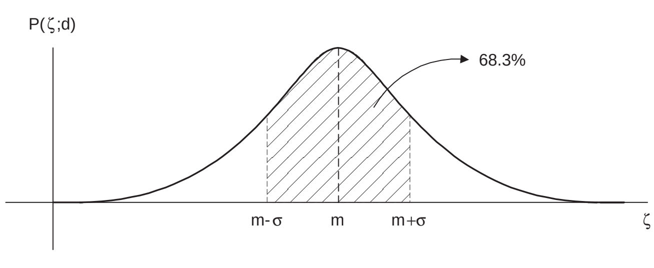
$$\int_{-\infty}^{\infty} \mathcal{P}(\zeta; d) d\zeta = \int_{-\infty}^{\infty} dF(\zeta; d) = 1 \tag{4.3}$$

In addition,

$$\int_a^b \mathcal{P}(\zeta; d)\ d\zeta = \int_a^b dF(\zeta; d) = F(b; d) - F(a; d) = Pr\{a < d \leq b\} \tag{4.4}$$

**Example:** Guassian or Normally Distributed Variable

$$\mathcal{P}(\zeta; d) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{\zeta - m}{\sigma}\right)^2\right\} \tag{4.5}$$




Note that this distribution is determined entirely by two parameters (the mean $m$ and standard deviation $\sigma$).

### 4.2.2 PROBABILITY DISTRIBUTION, DENSITY: VECTOR CASE

Let $d = \begin{bmatrix} d_1 & \cdots & d_n \end{bmatrix}^T$ be a *continuous* random variable vector($d \in \mathcal{R}^n$). Now we must quantify the distribution of its individual elements as well as their correlations.

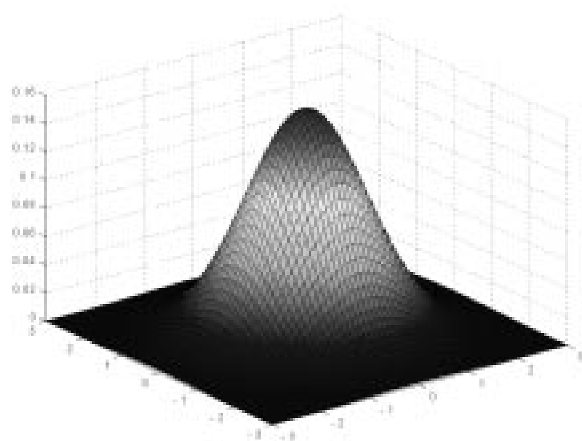
- **Joint Probability Distribution Function**
  The *joint probability distribution* function $F(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n)$ for random variable vector $d$ is defined as

$$F(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) = Pr\{d_1 \leq \zeta_1, \cdots, d_n \leq \zeta_n\} \tag{4.6}$$

  Now the domain of $F$ is an n-dimensional space. For example, for $n = 2$, $F$ is represented by a surface. Note that $F(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \to 1$ as $\zeta_1, \cdots, \zeta_n \to \infty$.

- **Joint and Marginal Probability Density Function**
  The *joint probability density* function $\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n)$ for random variable vector $d$ is defined as

$$\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) = \frac{\partial^n F(\zeta; d)}{\partial \zeta_1, \cdots, \zeta_n} \tag{4.7}$$



For convenience, we may write $\mathcal{P}(\zeta; d)$ to denote $\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n)$. Again,

$$\begin{aligned} & \textstyle\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \quad d\zeta_1 \cdots d\zeta_n \\ = \quad & \Pr\{a_1 < d_1 \leq b_1, \cdots, a_n < d_n \leq b_n\} \end{aligned} \tag{4.8}$$

Naturally,

$$\int_{-\infty}^{\infty}, \cdots, \int_{-\infty}^{\infty} \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) d\zeta_1 \cdots d\zeta_n = 1 \tag{4.9}$$

We can easily derive the probability density of individual element from the joint probability density. For instance,

$$\mathcal{P}(\zeta_1; d_1) = \int_{-\infty}^{\infty}, \cdots, \int_{-\infty}^{\infty} \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \quad d\zeta_2 \cdots d\zeta_n \tag{4.10}$$

This is called marginal probability density.

While the joint probability density (or distribution) tells us the likelihood of several random variables achieving certain values simultaneously, the marginal density tells us the likelihood of one element achieving certain value when the others are not known.

Note that in general

$$\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \neq \mathcal{P}(\zeta_1; d_1) \cdots \mathcal{P}(\zeta_n; d_n) \tag{4.11}$$

If

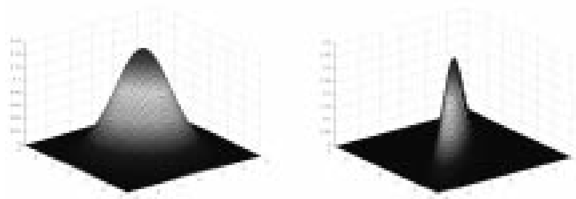
$$\mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) = \mathcal{P}(\zeta_1; d_1) \cdots \mathcal{P}(\zeta_n; d_n) \tag{4.12}$$

$d_1, \cdots, d_n$ are called *mutually independent*.

**Example:** Guassian or Jointly Normally Distributed Variables

Suppose that $\mathbf{d} \triangleq [d_1\ d_2]^T$ is a Gaussian variable. The density takes the form of

$$\begin{aligned}\mathcal{P}(\zeta_1, \zeta_2; d_1, d_2) &= \frac{1}{2\pi\sigma^1\sigma_2(1-\rho^2)^{1/2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{\zeta_1 - m_1}{\sigma_1}\right)^2\right.\right. \\ &\qquad \left.\left. -2\rho\frac{(\zeta_1 - m_1)(\zeta_2 - m_2)}{\sigma_1\sigma_2} + \left(\frac{\zeta_2 - m_2}{\sigma_2}\right)^2\right]\right\} \end{aligned} \tag{4.13}$$



Note that this density is determined by five parameters (the means $m_1, m_2$, standard deviations $\sigma_1, \sigma_2$ and correlation parameter $\rho$). $\rho = 1$ represents complete correlation between $d_1$ and $d_2$, while $\rho = 0$ represents no correlation.

It is fairly straightforward to verify that

$$\mathcal{P}(\zeta_1; d_1) = \int_{-\infty}^{\infty} \mathcal{P}(\zeta_1, \zeta_2; d_1, d_2) \quad d\zeta_2 \tag{4.14}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{1}{2}\left(\frac{\zeta_1 - m_1}{\sigma_1}\right)^2\right\} \tag{4.15}$$

$$\mathcal{P}(\zeta_2; d_2) = \int_{-\infty}^{\infty} \mathcal{P}(\zeta_1, \zeta_2; d_1, d_2) \quad d\zeta_1 \tag{4.16}$$

$$= \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{1}{2}\left(\frac{\zeta_2 - m_2}{\sigma_2}\right)^2\right\} \tag{4.17}$$

Hence, $(m_1, \sigma_1)$ and $(m_2, \sigma_2)$ represent parameters for the marginal density of $d_1$ and $d_2$ respectively. Note also that

$$\mathcal{P}(\zeta_1, \zeta_2; d_1, d_2) \neq \mathcal{P}(\zeta_1; d_1)\mathcal{P}(\zeta_2; d_2) \tag{4.18}$$

except when $\rho = 0$.

General n-dimensional Gaussian random variable vector $d = [d_1, \cdots, d_n]^T$ has the density function of the following form:

$$\mathcal{P}(\zeta; d) \triangleq \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \tag{4.19}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}} |P_d|^{1/2}} \exp\left\{ -\frac{1}{2}(\zeta - \bar{d})^T P_d^{-1} (\zeta - \bar{d}) \right\} \tag{4.20}$$

where the parameters are $\bar{d} \in \mathcal{R}^n$ and $P_d \in \mathcal{R}^{n \times n}$. The significance of these parameters will be discussed later.

### 4.2.3 EXPECTATION OF RANDOM VARIABLES AND RANDOM VARIABLE FUNCTIONS: SCALAR CASE

Random variables are completely characterized by their distribution functions or density functions. However, in general, these functions are nonparametric. Hence, random variables are often characterized by their moments up to a finite order; in particular, use of the first two moments is quite common.

- **Expection of Random Variable Fnction**
  Any function of $d$ is a random variable. Its expectation is computed as follows:
$$E\{f(d)\} \triangleq \int_{-\infty}^{\infty} f(\zeta)\mathcal{P}(\zeta; d) \quad d\zeta \tag{4.21}$$

- **Mean**
$$\bar{d} \triangleq E\{d\} = \int_{-\infty}^{\infty} \zeta \mathcal{P}(\zeta; d) \quad d\zeta \tag{4.22}$$
  The above is called mean or expectation of $d$.

- **Variance**

$$\mathrm{Var}\{d\} \triangleq E\{(d-\bar{d})^2\} = \int_{-\infty}^{\infty} (\zeta - \bar{d})^2 \mathcal{P}(\zeta; d) \;\; d\zeta \tag{4.23}$$

The above is the "variance" of $d$ and quantifies the extent of $d$ deviating from its mean.

**Example:** Gaussian Variable

For Gaussian variable with density

$$\mathcal{P}(\zeta; d) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \left( \frac{\zeta - m}{\sigma} \right)^2 \right\} \tag{4.24}$$

it is easy to verify that

$$\bar{d} \triangleq E\{d\} = \int_{-\infty}^{\infty} \zeta \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \left( \frac{\zeta - m}{\sigma} \right)^2 \right\} \;\; d\zeta = m \tag{4.25}$$

$$\mathrm{Var}\{d\} \triangleq E\{(d-\bar{d})^2\} = \int_{-\infty}^{\infty} (\zeta - m)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \left( \frac{\zeta - m}{\sigma} \right)^2 \right\} \;\; d\zeta = \sigma^2 \tag{4.26}$$

Hence, $m$ and $\sigma^2$ that parametrize the normal density represent the mean and the variance of the Gaussian variable.

### 4.2.4 EXPECTATION OF RANDOM VARIABLES AND RANDOM VARIABLE FUNCTIONS: VECTOR CASE

We can extend the concepts of mean and variance similarly to the vector case. Let $\mathbf{d}$ be a random variable vector that belongs to $\mathcal{R}^n$.

$$\begin{aligned} \bar{d}_\ell &= E\{d_\ell\} = \int_{-\infty}^{\infty} \zeta_\ell \mathcal{P}(\zeta_\ell; d_\ell) \;\; d\zeta_\ell \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \zeta_\ell \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \;\; d\zeta_1, \cdots, d\zeta_n \end{aligned} \tag{4.27}$$

$$\mathrm{Var}\{d_\ell\} = E\{(d_\ell - \bar{d}_\ell)^2\} = \int_{-\infty}^{\infty} (\zeta_\ell - \bar{d}_\ell)^2 \mathcal{P}(\zeta_\ell; d_\ell) \;\; d\zeta_\ell \tag{4.28}$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\zeta_\ell - \bar{d}_\ell)^2 \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \quad \zeta_1, \cdots, d\zeta_n$$

In the vector case, we also need to quantify the correlations among different elements.

$$\begin{aligned} \mathrm{Cov}\{d_\ell, d_m\} &= E\{(d_\ell - \bar{d}_\ell)(d_m - \bar{d}_m)\} \qquad (4.29) \\ = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} & (\zeta_\ell - \bar{d}_\ell)(\zeta_m - \bar{d}_m) \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \quad d\zeta_1, \cdots, d\zeta_n \end{aligned}$$

Note that

$$\mathrm{Cov}\{d_\ell, d_\ell\} = \mathrm{Var}\{d_\ell\} \qquad (4.30)$$

The ratio

$$\rho = \frac{\mathrm{Cov}\{d_\ell, d_m\}}{\sqrt{\mathrm{Var}\{d_\ell\}\mathrm{Var}\{d_m\}}} \qquad (4.31)$$

is the correlation factor. $\rho = 1$ indicates complete correlation ($d_\ell$ is determined uniquely by $d_m$ and vice versa). $\rho = 0$ indicates no correlation.

It is convenient to define covariance matrix for $d$, which contains all variances and covariances of $d_1, \cdots, d_n$.

$$\begin{aligned} \mathrm{Cov}\{d\} &= E\{(d - \bar{d})(d - \bar{d})^T\} \qquad (4.32) \\ = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} & (\zeta - \bar{d})(\zeta - \bar{d})^T \mathcal{P}(\zeta_1, \cdots, \zeta_n; d_1, \cdots, d_n) \quad d\zeta_1, \cdots, d\zeta_n \end{aligned}$$

The $(i, j)_{\mathrm{th}}$ element of $\mathrm{Cov}\{d\}$ is $\mathrm{Cov}\{d_i, d_j\}$. The diagonal elements of $\mathrm{Cov}\{d\}$ are variances of elements of $d$. The above matrix is symmetric since

$$\mathrm{Cov}\{d_i, d_j\} = \mathrm{Cov}\{d_j, d_i\} \qquad (4.33)$$

Covariance of two different vectors $x \in \mathcal{R}^n$ and $y \in \mathcal{R}^m$ can be defined similarly.

$$\mathrm{Cov}\{x, y\} = E\{(x - \bar{x})(y - \bar{y})^T\} \qquad (4.34)$$

In this case, Cov$\{x, y\}$ is an $n \times m$ matrix. In addition,

$$\mathrm{Cov}\{x, y\} = (\mathrm{Cov}\{y, x\})^T \tag{4.35}$$

**Example:** Gaussian Variables – 2-Dimensional Case
Let $d = [d_1 \ d_2]^T$ and

$$\begin{aligned} \mathcal{P}(\zeta; d) &= \frac{1}{2\pi\sigma^1\sigma_2(1-\rho^2)^{1/2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{\zeta_1 - m_1}{\sigma_1}\right)^2 \right.\right. \\ &\qquad \left.\left. -2\rho\frac{(\zeta_1 - m_1)(\zeta_2 - m_2)}{\sigma_1\sigma_2} + \left(\frac{\zeta_2 - m_2}{\sigma_2}\right)^2\right]\right\} \end{aligned} \tag{4.36}$$

Then,

$$\begin{aligned} E\{d\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \mathcal{P}(\zeta; d) \ \ d\zeta_1 d\zeta_2 \\ &= \begin{bmatrix} m_2 \\ m_2 \end{bmatrix} \end{aligned} \tag{4.37}$$

Similarly, one can show that

$$\begin{aligned} \mathrm{Cov}\{d\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \begin{bmatrix} \zeta_1 - m_1 \\ \zeta_2 - m_2 \end{bmatrix} \begin{bmatrix} (\zeta_1 - m_1) & (\zeta_2 - m_2) \end{bmatrix} \mathcal{P}(\zeta; d) \ \ d\zeta_1 d\zeta_2 \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \end{aligned} \tag{4.38}$$

**Example:** Gaussian Variables – n-Dimensional Case
Let $d = [d_1 \ \cdots \ d_n]^T$ and

$$\mathcal{P}(\zeta; d) = \frac{1}{(2\pi)^{\frac{n}{2}}|P_d|^{1/2}} \exp\left\{-\frac{1}{2}(\zeta - \bar{d})^T P_d^{-1}(\zeta - \bar{d})\right\} \tag{4.39}$$

Then, one can show that

$$
\begin{aligned}
E\{d\} &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \zeta \mathcal{P}(\zeta; d) \;\; d\zeta_1, \cdots, d\zeta_n = \bar{d} && (4.40) \\
\mathrm{Cov}\{d\} &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\zeta - \bar{d})(\zeta - \bar{d})^T \mathcal{P}(\zeta; d) \;\; d\zeta_1, \cdots, d\zeta_n = P_d && (4.41)
\end{aligned}
$$

Hence, $\bar{d}$ and $P_d$ that parametrize the normal density function $\mathcal{P}(\zeta; d)$ represent the mean and the covariance matrix.

**Exercise:** Verify that, with

$$
\bar{d} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}; \quad P_d = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix} \tag{4.42}
$$

one obtains the expression for normal density of a 2-dimensional vector shown earlier.

**NOTE: Use of SVD for Visualization of Normal Density**

Covariance matrix $P_d$ contains information about the spread (i.e., extent of deviation from the mean) for each element and their correlations. For instance,

$$
\begin{aligned}
\mathrm{Var}\{d_\ell\} &= [\mathrm{Cov}\{d\}]_{\ell,\ell} && (4.43) \\
\rho\{d_\ell, d_m\} &= \frac{[\mathrm{Cov}\{d\}]_{\ell,m}}{\sqrt{[\mathrm{Cov}\{d\}]_{\ell,\ell}\,[\mathrm{Cov}\{d\}]_{m,m}}} && (4.44)
\end{aligned}
$$

where $[\cdot]_{i,j}$ represents the $(i,j)_{\mathrm{th}}$ element of the matrix. However, one still has hard time understanding the correlations among all the elements and visualizing the overall shape of the density function. Here, the SVD can be useful. Because $P_d$ is a symmetric matrix, it has the following SVD:

$$
\begin{aligned}
P_d &\triangleq E\{(d - \bar{d})(d - \bar{d})^T\} && (4.45) \\
&= V\Sigma V^T && (4.46)
\end{aligned}
$$

$$= \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} \tag{4.47}$$

Pre-multiplying $V^T$ and post-multiplying $V$ to both sides, we obtain

$$E\{V^T(d-\bar{d})(d-\bar{d})^T V\} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \tag{4.48}$$

Let $d^* = V^T d$. Hence, $d^*$ is the representation of $d$ in terms of the coordinate system defined by orthonormal basis $v_1, \cdots, v_n$. Then, we see that

$$E\{(d^* - \bar{d}^*)(d^* - \bar{d}^*)^T\} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} \tag{4.49}$$

The diagonal covariance matrix means that every element of $d^*$ is completely independent of each other. Hence, $v_1, \cdots, v_n$ define the coordiate system with respect to which the random variable vector is independent. $\sigma_1^2, \cdots, \sigma_n^2$ are the variances of $d^*$ with respect to axes defined by $v_1, \cdots, v_n$.

**Exercise:** Suppose $d \in \mathcal{R}^2$ is zero-mean Gaussian and

$$P_d = \begin{bmatrix} 20.2 & 19.8 \\ 19.8 & 20.2 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \tag{4.50}$$

Then, $v_1 = [\frac{\sqrt{2}}{2} \; \frac{\sqrt{2}}{2}]^T$ and $v_2 = [\frac{\sqrt{2}}{2} \; -\frac{\sqrt{2}}{2}]^T$. Can you visualize the overall shape of the density function? What is the variance of $d$ along the (1,1) direction? What about along the (1,-1) direction? What do you think the conditional density of $d_1$ given $d_2 = \beta$ looks like? Plot the densities to verify.