Lecture 4: Probability Distributions and their Statistics

Lecturer: Prof. F. Jalayer, 12 February 2021
Notes prepared by: Prof. F. Jalayer and Dr. H. Ebrahimian

2. Parameters of the Probability Distribution

In general, the parameters (statistics) of a probability distribution can be classified in the following two general categories:

1. The moments of the distribution (e.g., mean, variance or standard deviation, and etc.) which are further broken down into two categories: *central* and *non-central* moments.
2. The ordered statistics (e.g., median, 16th and 84th percentiles) – This issue will be discussed in the next chapters.

2.1. *First moment of a probability distribution (non-central)*

The first moment of a probability distribution (also known as the *expected value*) for PDF of f_X and PMF of P is outlined in Eq. (31) and Eq. (32), respectively:

$$\mathbb{E}_X\left(x|\mathrm{I}\right)=\int_{\Omega_X} x\, f_X\left(x|\mathrm{I}\right)\mathrm{d}x \tag{31}$$

$$\mathbb{E}\left(X|\mathrm{I}\right)=\sum_{\forall x_i} x_i\, P\left(X=x_i|\mathrm{I}\right) \tag{32}$$

It is to note that $\mathbb{E}_X\left(x|\mathrm{I}\right)$ corresponding to a continuous distribution shows somehow the center of mass of the PDF; moreover, $\mathbb{E}\left(X|\mathrm{I}\right)$ for a discrete distribution does not necessarily need to be a possible value of x. For instance, for a discrete Uniform distribution, we have n possible value for x where a uniform PMF equal to $P(x|\mathrm{I})=1/n$ is assigned to each value of n; thus:

$$\mathbb{E}\left(X\middle|\mathrm{I}\right)=\sum_{i=1}^{n}x_i\,P\left(x_i\middle|\mathrm{I}\right)=\frac{\sum_{i=1}^{n}x_i}{n} \tag{33}$$

For a continuous Uniform distribution, we have,

$$\mathbb{E}_X\left(x\middle|\mathrm{I}\right)=\int_a^b x\cdot\frac{1}{b-a}\mathrm{d}x=\frac{b^2-a^2}{2\left(b-a\right)}=\frac{a+b}{2} \tag{34}$$

It is to note that the mean of our data might not be equal to any x. Therefore, in case of having equi-probable data, the expected value coincides with the mean. Another key feature of "$\mathbb{E}$" is that it is a *linear operator*, i.e.:

$$\mathbb{E}\left(\sum_{i=1}^{k}\left(a_iX_i+b_i\right)\middle|\mathrm{I}\right)=\sum_{i=1}^{k}\left[a_i\mathbb{E}\left(X_i\middle|\mathrm{I}\right)+b_i\right] \tag{35}$$

2.2. *Higher moments of a probability distribution*

The m^{th} non-central moment of a distribution can be defined as (see also Eq. 31 and Eq. 32):

$$\mathbb{E}_X\left(x^m\middle|\mathrm{I}\right)\equiv\int_{\Omega_X}x^m\,f_X\left(x\middle|\mathrm{I}\right)\mathrm{d}x \tag{36}$$

$$\mathbb{E}\left(X^m\middle|\mathrm{I}\right)=\sum_{\forall x_i}x_i^m\,P\left(x_i\middle|\mathrm{I}\right) \tag{37}$$

Hence, the zero-moment (i.e., m=0) is equal to 1.

2.3. *Central moments of a probability distribution*

The m^{th} central moment of a probability distribution (continuous or discrete) is defined as:

$$\mathbb{E}_X\left(\left[x-\mathbb{E}_X\left(x\middle|\mathrm{I}\right)\right]^m\middle|\mathrm{I}\right)\equiv\int_{\Omega_X}\left[x-\mathbb{E}_X\left(x\middle|\mathrm{I}\right)\right]^m f_X\left(x\middle|\mathrm{I}\right)\mathrm{d}x \tag{38}$$

$$\mathbb{E}\left(\left[X-\mathbb{E}\left(X|\mathrm{I}\right)\right]^{m}|\mathrm{I}\right)\equiv\sum_{\forall x_i}\left[x_i-\mathbb{E}\left(X|\mathrm{I}\right)\right]^{m}P\left(x_i|\mathrm{I}\right) \tag{39}$$

The term $x-\mathbb{E}\left(X|\mathrm{I}\right)$ illustrates the residuals of x around the expected value. For m=0 and 1, trivial solutions are obtained. However, the case of m=2, i.e. the second central moment, is called the *variance* of a distribution, which can be estimated as:

$$\begin{aligned}\mathbb{VAR}_X\left(x|\mathrm{I}\right)&\equiv\sigma^2_{X|\mathrm{I}}=\mathbb{E}_X\left(\left[x-\mathbb{E}_X\left(x|\mathrm{I}\right)\right]^2|\mathrm{I}\right)=\mathbb{E}_X\left(x^2-2x\mathbb{E}_X\left(x|\mathrm{I}\right)+\left[\mathbb{E}_X\left(x|\mathrm{I}\right)\right]^2|\mathrm{I}\right)\\&=\mathbb{E}_X\left(x^2|\mathrm{I}\right)-2\left[\mathbb{E}_X\left(x|\mathrm{I}\right)\right]^2+\left[\mathbb{E}_X\left(x|\mathrm{I}\right)\right]^2\\&=\underbrace{\mathbb{E}_X\left(x^2|\mathrm{I}\right)}_{\text{second non-central moment}}-\underbrace{\left[\mathbb{E}_X\left(x|\mathrm{I}\right)\right]^2}_{\text{first non-central moment}}\end{aligned} \tag{40}$$

The parameter σ is called the standard deviation of the distribution. A key feature of the variance is that it is not a *linear operator*. Let us verify:

$$\begin{aligned}\mathbb{VAR}\left(aX+bY|\mathrm{I}\right)&=\mathbb{E}\left[\left(aX+bY\right)^2|\mathrm{I}\right]-\left[\mathbb{E}\left(aX+bY|\mathrm{I}\right)\right]^2\\&=a^2\mathbb{VAR}\left(X|\mathrm{I}\right)+b^2\mathbb{VAR}\left(Y|\mathrm{I}\right)+2ab\left[\mathbb{E}\left(XY|\mathrm{I}\right)-\mathbb{E}\left(X|\mathrm{I}\right)\mathbb{E}\left(Y|\mathrm{I}\right)\right]\\&=a^2\mathbb{VAR}\left(X|\mathrm{I}\right)+b^2\mathbb{VAR}\left(Y|\mathrm{I}\right)+2ab\,\mathbb{COV}\left(X,Y|\mathrm{I}\right)\\&=a^2\sigma^2_{X|\mathrm{I}}+b^2\sigma^2_{Y|\mathrm{I}}+2ab\rho_{XY|\mathrm{I}}\cdot\sigma_{X|\mathrm{I}}\cdot\sigma_{Y|\mathrm{I}}\end{aligned} \tag{41}$$

where $\mathbb{COV}\left(X,Y|\mathrm{I}\right)$ is denoted as the *covariance* of the two random variables, and $\rho_{XY|\mathrm{I}}$ is called the *coefficient of correlation* and is obtained as:

$$\rho_{XY|\mathrm{I}}=\frac{\mathbb{COV}\left(X,Y|\mathrm{I}\right)}{\sigma_{X|\mathrm{I}}\,\sigma_{Y|\mathrm{I}}}=\frac{\mathbb{E}\left(XY|\mathrm{I}\right)-\mathbb{E}\left(X|\mathrm{I}\right)\mathbb{E}\left(Y|\mathrm{I}\right)}{\sigma_{X|\mathrm{I}}\,\sigma_{Y|\mathrm{I}}} \tag{42}$$

Actually, the last term in Eq. (41) defines the possible correlation between the two uncertain parameters X, and Y. In case that $\rho_{XY|\mathrm{I}}=0$ (or the covariance is zero), one can easily obtain that (according to Eq. 38):

$$\mathbb{E}\left(XY\left|\mathrm{I}\right.\right)=\mathbb{E}\left(X\left|\mathrm{I}\right.\right)\mathbb{E}\left(Y\left|\mathrm{I}\right.\right) \tag{43}$$

Now, assuming that the two uncertain parameters X and Y are independent; hence,

$$\begin{aligned}&\mathbb{E}\left(XY\left|\mathrm{I}\right.\right)=\int_{\Omega_X}\int_{\Omega_Y} xyf_{XY}\left(x,y\left|\mathrm{I}\right.\right)\mathrm{d}x\,\mathrm{d}y=\int_{\Omega_X} x\,f_X\left(x\left|\mathrm{I}\right.\right)\mathrm{d}x\cdot\int_{\Omega_Y} y\,f_Y\left(y\left|\mathrm{I}\right.\right)\mathrm{d}y=\mathbb{E}\left(X\left|\mathrm{I}\right.\right)\mathbb{E}\left(Y\left|\mathrm{I}\right.\right)\\&\Leftrightarrow \mathbb{COV}\left(X,Y\left|\mathrm{I}\right.\right)=0\end{aligned} \tag{44}$$

This reveals that if the two uncertain parameters are independent (which relates to the whole distribution), their covariance equals zero and hence, they are uncorrelated. However, the reverse condition is not always true; i.e., Eq. (43) cannot lead to the conclusion of Eq. (29) (since Eq. 43 does not reveal anything about the whole distribution or the higher moments).

It is also noteworthy that the interval $-1.0\le\rho_{XY|\mathrm{I}}\le 1.0$ always holds. To provide a proof for this fact, re-write Eq. (36) as follows taking into account the fact that variance is always positive:

$$\mathbb{VAR}\left(aX+bY\left|\mathrm{I}\right.\right)=\left(\sigma^2_{X|\mathrm{I}}\right)a^2+\left[2b\rho_{XY|\mathrm{I}}\,\sigma_{X|\mathrm{I}}\,\sigma_{Y|\mathrm{I}}\right]a+b^2\sigma^2_{Y|\mathrm{I}}\ge 0 \tag{45}$$

Therefore, for the quadratic expression (with respect to a) in Eq. (45) to be positive, the following expression should hold:

$$\left[b\rho_{XY|\mathrm{I}}\,\sigma_{X|\mathrm{I}}\,\sigma_{Y|\mathrm{I}}\right]^2-b^2\sigma^2_{X|\mathrm{I}}\,\sigma^2_{Y|\mathrm{I}}\le 0\Rightarrow\rho^2_{XY|\mathrm{I}}\le 1\Rightarrow -1.0\le\rho_{XY|\mathrm{I}}\le 1.0 \tag{46}$$

In case that $\rho_{XY|\mathrm{I}}=1$, the variables are *fully positively correlated*, while in case of $\rho_{XY|\mathrm{I}}=-1$, they are *fully negatively correlated*.

2.4. *Parameters of Bernoulli distribution*

According to Eq. (31), the expected value is:

$$\mathbb{E}\left(X\left|\mathrm{I}\right.\right)=\sum_{i=1}^{2}x_i\,P\left(x_i\left|\mathrm{I}\right.\right)=0\cdot\left(1-\pi\right)+1\cdot\pi=\pi \tag{47}$$

Based on Eq. (40), the variance is:

$$\mathbb{VAR}\left(X\left|\mathrm{I}\right.\right) \equiv \sigma^2_{X|\mathrm{I}} = \mathbb{E}\left(X^2\left|\mathrm{I}\right.\right) - \left[\mathbb{E}\left(X\left|\mathrm{I}\right.\right)\right]^2 = \pi - \pi^2 = \pi\left(1-\pi\right) \tag{48}$$

The expected value is equal to the probability of success and the variance is the probability of success multiplied by probability of failure.

2.5. *Parameters of Binomial distribution*

The number of successes r can be actually interpreted as the sum of n **i**ndependent and **i**dentically **d**istributed (*i.i.d*) Binary (Bernoulli) variables X_i, i=1:n. Hence, based on Eq. (31) and Eq. (47), the expected value will become

$$\mathbb{E}\left(r\left|n,\pi\right.\right) = \sum_{i=1}^{n}\left[\mathbb{E}\left(X_i\left|\mathrm{I}\right.\right)\right] = \sum_{i=1}^{n}\pi = n\pi \tag{49}$$

Accordingly, based on Eq. (40) and Eq. (48), the variance will become:

$$\mathbb{VAR}\left(r\left|n,\pi\right.\right) = \sum_{i=1}^{n}\left[\mathbb{VAR}\left(X_i\left|\mathrm{I}\right.\right)\right] = n\pi\left(1-\pi\right) \tag{50}$$

Note that since variables X_i are independent, they are also uncorrelated (the correlation coefficient for any distinct pair of X_i values is equal to 0). Therefore the variance of the sum of Xi's can be calculated as the sum of the variances.