(Univariate) distribution of X: Example:

$X =$	0	1	2	3
Pr:	$\frac{3}{10}$	$\frac{4}{10}$	$\frac{2}{10}$	$\frac{1}{10}$

$$
\begin{aligned}
\mu &= 0 \cdot \frac{3}{10} + 1 \cdot \frac{4}{10} + 2 \cdot \frac{2}{10} + 3 \cdot \frac{1}{10} = \frac{11}{10} \\
\mathrm{Var}(X) &= \mathbb{E}\left((X-\mu)^2\right) = \mathbb{E}\left(X^2\right) - \mu^2 = \frac{21}{10} - \left(\frac{11}{10}\right)^2 = \frac{89}{100} \\
&\Rightarrow \sigma = \sqrt{0.89} = 0.9434 \\
\text{Skewness} &= \frac{\mathbb{E}\left((X-\mu)^3\right)}{\sigma^3} \qquad \text{(dimensionless)} \\
\text{Kurtosis} &= \frac{\mathbb{E}\left((X-\mu)^4\right)}{\sigma^4}
\end{aligned}
$$

Transforming X

$$Y = (X-2)^2$$

defines a new RV Y with its own (not so easy to construct - topic of Chapter 9) distribution. BUT, it is still easy to compute its expected value (and variance), thus

$$
\begin{aligned}
\mathbb{E}(Y) &= (0-2)^2 \cdot \frac{3}{10} + (1-2)^2 \cdot \frac{4}{10} + (2-2)^2 \cdot \frac{2}{10} + (3-2)^2 \cdot \frac{1}{10} = \frac{17}{10} \\
\mathbb{E}(g(X)) &= \sum_{\text{All } i} g(i) \cdot \Pr(X=i)
\end{aligned}
$$

Note that the answer is NOT equal to $(\mu-2)^2$!!!

In general

$$\mathbb{E}(g(X)) = \sum_{\text{All } i} g(i) \cdot \Pr(X=i) \neq g(\mu)$$

There is one important exception to this, when the transformation is **linear**, meaning $Y = a \cdot X + b$, in which case

$$
\begin{aligned}
\mathbb{E}(Y) &= a \cdot \mu + b \\
\mathrm{Var}(X) &= a^2 \cdot \mathrm{Var}(X)
\end{aligned}
$$

Bivariate case: We have to be able to construct two *marginal* distributions and any one of the *conditional* distributions, eg.

$X \mid Y = -1$	1	2	3
Pr:	$\frac{12}{33}$	$\frac{13}{33}$	$\frac{8}{33}$

base on which we can compute $\mathbb{E}(X \mid Y = -1)$, $\mathrm{Var}(X \mid Y = -1)$ etc.

Bivariate transformation:

$$U = h(X, Y)$$

Not so easy to find the new (univariate) distribution of U, but easy enough to find the moments of U, e.g. $U = (X - Y)^2$ by building a table of the values of U, thus

1	4	9	16
0	1	4	9
1	0	1	4

then multiplying each of these values by the corresponding (bivariate) probability and adding the results over the whole table, getting

$$\mathbb{E}(U) = \frac{4 \cdot 12 + 9 \cdot 13 + 16 \cdot 8 + 15 + 9 \cdot 10 + 9 + 6}{100} = 4.13$$

In general:

$$\mathbb{E}(h(X,Y)) = \sum_{\text{All } i,j} h(i,j) \cdot \Pr(X = i \cap Y = j) \neq h(\mu_x, \mu_y)$$

Important **exception**, when the (bivariate) transformation is **linear**, ie. $U = a \cdot X + b \cdot Y + c$

$$\begin{aligned} \mathbb{E}(a \cdot X + b \cdot Y + c) &= a \cdot \mu_x + b \cdot \mu_y + c \\ \mathrm{Var}(a \cdot X + b \cdot Y + c) &= a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y) \\ &\quad + 2ab \cdot \mathrm{Cov}(X,Y) \end{aligned}$$

Covariance of X and Y is defined as their $(1^{\mathrm{st}}, 1^{\mathrm{st}})$ joint central moment, namely

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left((X - \mu_x) \cdot (Y - \mu_y)\right)$$

Independence of X and Y (easy to tell) implies that

$$\Pr(X = i \cap Y = j) = \Pr(X = i) \cdot \Pr(Y = j)$$

for any i, j pair, implying that we no longer need a table of their joint probabilities - all we need is the two individual (univariate) distributions. Independence implies that in general

$$\mathbb{E}(g(X) \cdot h(Y)) = \mathbb{E}(g(X)) \cdot \mathbb{E}(h(Y))$$

Special case: $\mathrm{Cov}(X,Y) = 0$ (but not reverse)! This simplifies the formula for Var of linear combination (of any number of independent RVs).

Correlation coefficient is defined by

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_x \cdot \sigma_y} \qquad \text{also dimensionless}$$

(equal to zero when X and Y are independent). One can show that $-1 \leq \rho \leq 1$.

Final (important) definition of *probability generating function* (PGF) of a univariate distribution:

$$P(z) = \mathbb{E}\left(z^X\right) = p_0 + p_1 z + p_2 z^2 + p_3 z^3 + \ldots$$

enables us to compute (rather easily) the **factorial** moments of X:

$$\begin{aligned}
\mathbb{E}(X) &= P'(z)|_{z=1} \\
\mathbb{E}(X(X-1)) &= P''(z)|_{z=1} \\
\mathbb{E}(X(X-1)(X-2)) &= P'''(z)|_{z=1}
\end{aligned}$$

When X and Y are *independent*

$$P_{X+Y}(z) = P_x(z) \cdot P_y(z)$$

Special case: when X_1, $X_2, ...X_n$ are IID (called **random independent sample**)

$$P_{\text{sum}}(z) = P(z)^n$$

Central Limit Theorem (aka Normal approximation) claims that the corresponding $\sum_{i=1}^{n} X_i$ (a single RV) has the mean of $n\mu$ and variance of $n\sigma^2$ (which follows from the above rules), and its distribution is approximately **Normal** (with the same mean and variance). This implies that

$$\Pr\left(\sum_{i=1}^{n} X_i \leq c\right) \simeq \Pr\left(Z < \frac{\tilde{c} - n\mu}{\sqrt{n} \cdot \sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\text{rhs}} \exp\left(-\frac{z^2}{2}\right) dz$$

where Z is a RV having the **standardized** (mean 0, variance equal to 1) Normal distribution. When sampling a *discrete* distribution, c must be slightly increased (to $\tilde{c}$) to lie exactly half way between two potential values of the sum; this is the so-called **continuity correction**.