EXPECTED VALUE of a RV

corresponds to the *average* value one would get for the RV when repeating the experiment, *independently*, infinitely many times. More accurately, consider a `Random Independent Sample (RIS)` of n values of X (e.g. 0, 1, 1, 0, 2, 1, 2, 1, 1, 0) and the corresponding `Sample Mean` $\bar{X} \equiv \frac{\sum_{j=1}^{n} X_j}{n} \equiv \sum_{\text{All } i} i \times f_i \ \left(= 0 \times \frac{3}{10} + 1 \times \frac{5}{10} + 2 \times \frac{2}{10} = 0.9\right)$, where f_i are the `observed` ***`relative`*** `frequencies` (of each possible value). We know that, as n (the `sample size`) increases, each f_i tends to the corresponding $\Pr(X = i)$.This implies that, in the same limit, $\bar{X}$ tends to $\sum_{\text{All } i} i \cdot \Pr(X = i)$, which we denote $\mathbb{E}(X)$ and call the `expected (mean)` value of X (so, in the end, the expected value is computed from the theoretical probabilities, not by doing the experiment). Note that $\sum_{\text{All } i} i \cdot \Pr(X = i)$ is a 'weighted' average of all possible values of X by their probabilities. And, as we have two names for it, we also have two alternate notations, $\mathbb{E}(X)$ is sometimes denoted μ_x (the Greek 'mu').

EXAMPLES:

- When X is the number of dots in a single roll of a die, we get $\mathbb{E}(X) = \frac{1+2+3+4+5+6}{6} = 3.5$ (for a symmetric distribution, the mean is at the centre of symmetry - for any other distribution, it is at the 'center of mass'). Note that the name 'expected' value is somehow misleading.

- Let Y be the larger of the two numbers when rolling two dice. $\mathbb{E}(Y) = 1 \times \frac{1}{36} + 2 \times \frac{3}{36} + 3 \times \frac{5}{36} + 4 \times \frac{7}{36} + 5 \times \frac{9}{36} + 6 \times \frac{11}{36} = \frac{1+6+15+28+45+66}{36} = \frac{161}{36} = 4.47\bar{2}$.

- Let U have the following (arbitrarily chosen) distribution:

$U =$	0	1	2	3	4
Prob:	0.3	0.2	0.4	0	0.1

  $\mathbb{E}(U) = 0.2 + 0.8 + 0.4 = 1.4$.

What happens to the expected value of X when we **transform** it, i.e. define a new RV by: $U \equiv \frac{X}{1+X}$, or $g(X)$ in general?

The main thing to remember is that, *in general*, the mean does *not* transform accordingly, i.e. $\mu_u \neq \frac{\mu_x}{1+\mu_x}$, etc. This is also true for a transformation of X and Y, i.e. $\mathbb{E}[h(X,Y)] \neq h(\mu_x, \mu_y)$.

But (at least one good news), to find the expected value of $g(X)$, $h(X,Y)$, ..., we can bypass constructing the new distribution (which was a tedious process) and use:

$$\begin{aligned}\mathbb{E}[g(X)] &= \sum_{\text{All } i} g(i) \times \Pr(X = i) \\ \mathbb{E}[h(X,Y)] &= \sum_{\text{All } i,j} h(i,j) \times \Pr(X = i \cap Y = j)\end{aligned}$$

EXAMPLE:

Based on U, we define $W = |U-2|^3$ and compute $\mathbb{E}(U) = 8\times0.3+1\times0.2+0\times0.4+8\times0.1 = 3.4$. Clearly, it would have been a mistake to use $|1.4-2|^3 = 0.216$ (totally wrong).

We can also get the correct answer the 'long' way (just to verify the 'short' answer), by first finding the distribution of W:

$W =$	8	1	0	8
$U =$	0	1	2	4
Prob:	0.3	0.2	0.4	0.1

implying

$W =$	0	1	8
Prob:	0.4	0.2	0.4

. Based on this, $\mathbb{E}(W) = 0.2 + 3.2 = 3.4$ (check).

Exception: **Linear Transformations**

Only for these, we can find the mean of the new RV by simply replacing X by μ_x, thus:

$$\mathbb{E}(a \cdot X + c) = a \cdot \mu_x + c$$

Proof: $\mathbb{E}(a \cdot X + c) = \sum_{\text{All } i} (a \cdot i + c) f_x(i) = a \sum_{\text{All } i} i \times f_x(i) + c \sum_{\text{All } i} f_x(i) = a \cdot \mu_x + c$ ■

EXAMPLE: $\mathbb{E}(2U - 3) = 2 \times 1.4 - 3 = -0.2$

Expected values related to a bivariate distribution

When a bivariate distribution is given, the easiest way to compute the *individual* expected values (of X and Y) is through the marginals.

EXAMPLE:

	$X =$ 1	2	3	
$Y =$ 0	0.1	0	0.3	0.4
1	0.3	0.1	0.2	0.6
	0.4	0.1	0.5	

we compute $\mathbb{E}(X) = 1 \times 0.4 + 2 \times 0.1 + 3 \times 0.5 = 2.1$ and $\mathbb{E}(Y) = 0 \times 0.4 + 1 \times 0.6 = 0.6$.

This is how we also deal with an expected value of $g_1(X)$ and/or $g_2(Y)$.

Only when the new variable is defined by $h(X, Y)$, we have 'weigh-average' the whole table. For example: $\mathbb{E}(X \cdot Y) = 1 \times 0 \times 0.1 + 2 \times 0 \times 0 + 3 \times 0 \times 0.3 + 1 \times 1 \times 0.3 + 2 \times 1 \times 0.1 + 3 \times 1 \times 0.2 = 1.1$

More EXAMPLES:

- $\mathbb{E}\left[(X-1)^2\right] = 0^2 \times 0.4 + 1^2 \times 0.1 + 2^2 \times 0.5 = 2.1$
- $\mathbb{E}\left[\frac{1}{1+Y^2}\right] = \frac{1}{1+0^2} \times 0.4 + \frac{1}{1+1^2} \times 0.6 = 0.7$
- $\mathbb{E}\left[\frac{(X-1)^2}{1+Y^2}\right]$ (multiplying the last two results would be *wrong*). Here it may help to first build the corresponding table of the $\frac{(X-1)^2}{1+Y^2}$ values: $\begin{array}{|c|c|c|}\hline 0 & 1 & 4 \\ \hline 0 & \frac{1}{2} & 2 \\ \hline \end{array}$. Answer: $1.2 + 0.05 + 0.4 = 1.65$

For **Linear Transformation** (of *two* RVs) we get:

$$\mathbb{E}(a \cdot X + b \cdot Y + c) = a \cdot \mu_x + b \cdot \mu_y + c$$

Proof: $\mathbb{E}(aX + bY + c) = \sum_i \sum_j (a \times i + b \times j + c) f_{xy}(i, j) = a \sum_i i \times f_x(i) + b \sum_j j \times f_y(j) + c = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y) + c$ Note that X and Y need *not* be *independent*!

EXAMPLE: Using the previous bivariate distribution, $\mathbb{E}(2X - 3Y + 4)$ is simply

$2 \times 2.1 - 3 \times 0.6 + 4 = 6.4$

The previous formula easily extends to *any number* of RVs (again, not necessarily independent!)

$$\mathbb{E}(a_1 X_1 + a_2 X_2 + ... + a_k X_k + c) =$$

$$a_1 \mathbb{E}(X_1) + a_2 \mathbb{E}(X_2) + ... + a_k \mathbb{E}(X_k) + c$$

■

Can **Independence** help (in other cases of expected value)?

Yes, the expected value of a *product* of RVs equals the product of the individual expected values, but ONLY when these RVs are *independent*:

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

Proof:

$\mathbb{E}(X \cdot Y) = \sum_i \sum_j i \times j \times f_x(i) \times f_y(j) = \left(\sum_i i \times f_x(i) \right) \times \left(\sum_j j \times f_y(j) \right) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$ ■

The statement can actually be made more **general**: When X and Y are *independent*

$$\mathbb{E}\left[g_1(X) \cdot g_2(Y)\right] = \mathbb{E}\left[g_1(X)\right] \cdot \mathbb{E}\left[g_2(Y)\right]$$

Moments of a RV

There are two types of 'moments', `Simple Moments`

$$\mathbb{E}(X^k)$$

and `Central Moments`

$$\mathbb{E}\left[(X-\mu_x)^k\right]$$

where k is an integer.

Special cases: The 0^{th} moment is identically equal to 1. The first *simple* moment is μ_x (yet another name for it!). The second simple moment is $\mathbb{E}(X^2) \neq \mu_x^2$, etc. The first *central* moment is identically equal to 0. The second central moment $\mathbb{E}\left[(X-\mu_x)^2\right]$ must be ≥ 0 (averaging non-negative quantities cannot result in a negative number). It is of such importance that it gets its own name: the **variance** of X, denoted $\text{Var}(X)$. When doing the computation 'by hand', it helps to realize that $\mathbb{E}\left[(X-\mu_x)^2\right] = \mathbb{E}(X^2) - \mu_x^2$.

As a measure of the *spread* of the distribution of X values, we take $\sigma_x \equiv \sqrt{\text{Var}(X)}$ and call it the **standard deviation** of X. For all distributions, the $(\mu - \sigma,\ \rho + \sigma)$ interval should contain the 'bulk' of the distribution, i.e. anywhere from 50 to 90% (in terms of the corresponding histogram).

Finally, **skewness** is defined as $\dfrac{\mathbb{E}\left[(X-\mu_x)^3\right]}{\sigma_x^3}$ (it measures to what extent is the distribution non-symmetric, or 'skewed'), and **kurtosis** as $\dfrac{\mathbb{E}\left[(X-\mu_x)^4\right]}{\sigma_x^4}$ (it measures the degree of 'flatness', 3 being its typical value, higher for 'peaked', smaller for 'flat' distributions). These are a lot less important than the mean and variance (later, we will understand why).

EXAMPLES:

- X is the number of dots when rolling one die: $\mu_x = \frac{7}{2}$, $\text{Var}(X) = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$ implying $\sigma_x = \sqrt{\frac{35}{12}} = 1.7078$. Note that 3.5 ± 1.708 contains 66.7% of the distribution. Skewness, for a symmetric distribution, is always equal to 0, kurtosis can be computed from $\mathbb{E}[(X-\mu)^4] = \frac{(-2.5)^4+(-1.5)^4+(-0.5)^4+0.5^4+1.5^4+2.5^4}{6} = 14.729 \Rightarrow$ kurtosis $= \dfrac{14.729}{(\frac{35}{12})^2} = 1.7314$ ('flat').

- For

$U =$	0	1	2	4
Prob:	0.3	0.2	0.4	0.1

  $\mu_u = 1.4$, $\text{Var}(U) = 0^2 \times 0.3 + 1^2 \times 0.2 + 2^2 \times 0.4 + 4^2 \times 0.1 - 1.4^2 = 1.44 \Rightarrow \sigma_u = \sqrt{1.44} = 1.2$. From $\mathbb{E}[(U-\mu_u)^3] = (-1.4)^3 \times 0.3 + (-0.4)^3 \times 0.2 + 0.6^3 \times 0.4 + 2.6^3 \times 0.1 = 1.008$, the skewness is $\dfrac{1.008}{1.2^3} = 0.58\bar{3}$ (long right tail), and from $\mathbb{E}[(U-\mu_u)^4] = (-1.4)^4 \times 0.3 + (-0.4)^4 \times 0.2 + 0.6^4 \times 0.4 + 2.6^4 \times 0.1 = 5.7792$ the kurtosis equals $\frac{5.7792}{1.2^4} = 2.787$

When X is transformed to $Y = g(X)$, we already know that there is no general 'shortcut' for computing $\mathbb{E}(Y)$. This (even more so) applies to the *variance* of Y, which also needs to be computed 'from scratch'. But, we did manage to simplify the expected value of a **linear transformation** of X (i.e., of $Y = aX + c$). Is there any direct conversion of $\text{Var}(X)$ into $\text{Var}(Y)$ in this (linear) case?

The answer is 'yes', and we can easily derive the corresponding formula: $\text{Var}(aX+c) = \mathbb{E}[(aX+c)^2] - (a\mu_x + c)^2 = \mathbb{E}[a^2X^2 + 2acX + c^2] - (a\mu_x + c)^2 = a^2\mathbb{E}(X^2) - a^2\mu_x^2 = a^2\text{Var}(X)$. This implies that

$$\sigma_{aX+c} = |a| \cdot \sigma_X$$

Moments - the bivariate case

Firstly, there will be the individual moments of X (and, separately, Y), which can be established based on the corresponding *marginal* distribution.

Are there any other `(joint) moments`? Yes, and again we have the *simple* moments $\mathbb{E}(X^k \cdot Y^m)$ and the *central* moments $\mathbb{E}\left[(X-\mu_x)^k \cdot (Y-\mu_y)^m\right]$. The most important of these is the *first,first central* moment called the `covariance` of X and Y:

$$\text{Cov}(X,Y) = \mathbb{E}\left[(X-\mu_x)\cdot(Y-\mu_y)\right] \equiv \mathbb{E}(X\cdot Y) - \mu_x\cdot\mu_y$$

It is obviously 'symmetric', i.e. Cov(X,Y) =Cov(Y,X) and it becomes *zero* when X and Y are *independent* (but not necessarily the other way round).

Based on Cov(X,Y), one can define the `Correlation Coefficient` between X and Y by:

$$\rho_{xy} = \frac{\text{Cov}(X,Y)}{\sigma_x\cdot\sigma_y}$$

(Greek letter 'rho'). Its value must be always between -1 and 1.

Proof: Obviously,

$$\text{Var}(X-bY) = \text{Var}(X) - 2b\text{Cov}(X,Y) + b^2\text{Var}(Y) \geq 0$$

for *any* value of b, including

$$b = \tfrac{\text{Cov}(X,Y)}{\text{Var}(Y)}$$

This implies that

$$\begin{aligned}&\text{Var}(X) - 2\tfrac{\text{Cov}(X,Y)}{\text{Var}(Y)}\text{Cov}(X,Y) + \tfrac{\text{Cov}(X,Y)^2}{\text{Var}(Y)^2}\text{Var}(Y)\\ &= \text{Var}(X) - \tfrac{\text{Cov}(X,Y)^2}{\text{Var}(Y)} \geq 0\end{aligned}$$

which, after dividing by Var(X) yields

$$1 \geq \rho^2$$

■

EXAMPLE: For one of our previous distributions

$Y =$ \ $X =$	1	2	3	
0	0.1	0	0.3	0.4
1	0.3	0.1	0.2	0.6
	0.4	0.1	0.5	

we get $\mu_x = 2.1$, $\mu_y = 0.6$, $\text{Var}(X) = 5.3 - 2.1^2 = 0.89$, $\text{Var}(Y) = 0.6 - 0.6^2 = 0.24$, $\text{Cov}(X, Y) = 0.3 + 0.2 + 0.6 - 2.1 \times 0.6 = -0.16$ (may be negative), and $\rho_{xy} = \frac{-0.16}{\sqrt{0.89 \times 0.24}} = -0.3462$ ■

One can also show that $\rho_{aX+c,bY+d} = \frac{\text{Cov}(aX+c,bY+d)}{\sigma_{aX+c} \cdot \sigma_{bY+d}} = \frac{a \cdot b \cdot \text{Cov}(X,Y)}{|a| \cdot |b| \cdot \sigma_X \cdot \sigma_Y} = \pm\rho_{xy}$ (linear transformation does not change the value of ρ, but it may change its sign - can you tell when?).

Linear Combination of RVs

Starting with X and Y, let's see whether we can simplify the expression for $\text{Var}(aX + bY + c) = \mathbb{E}\left[(aX + bY + c)^2\right] - \left(a\mu_x + b\mu_y + c\right)^2 = a^2\mathbb{E}(X^2) + b^2\mathbb{E}(Y^2) + 2ab\mathbb{E}(X \cdot Y) - a^2\mu_x^2 - b^2\mu_x^2 - 2ab\mu_x\mu_y =$

$$a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

Independence eliminates the last term.

This result can be easily **extended** to a linear combination of *any number* of random variables:

$$\text{Var}(a_1X_1 + a_2X_2 + ...a_kX_k + c) =$$

$$a_1^2\text{Var}(X_1) + a_2^2\text{Var}(X_2) + ... + a_k^2\text{Var}(X_k) +$$

$$2a_1a_2\text{Cov}(X_1, X_2) + 2a_1a_3\text{Cov}(X_1, X_3) + ...$$

$$... + 2a_{k-1}a_k\text{Cov}(X_{k-1}, X_k)$$

Mutual *independence* (if present) eliminates the last row of $\binom{k}{2}$ *covariances.*

And finally a formula for a covariance of one *linear combination* of RVs against another:

$$\text{Cov}\left(a_1X_1 + a_2X_2 + ..., b_1Y_1 + b_2Y_2 + ...\right) =$$

$$a_1b_1\text{Cov}(X_1, Y_1) + a_1b_2\text{Cov}(X_1, Y_2)$$

$$+a_2b_1\text{Cov}(X_2, Y_1) + a_2b_2\text{Cov}(X_2, Y_2) + ...$$

(I will call this the *distributive law* of covariance).

Sample Mean and its distribution

Consider a random independent sample of size n from an arbitrary distribution. We know that the `sample mean`

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

is itself a RV with its own distribution. Regardless how that distribution looks like, its expected value must be the same as the expected value of the distribution from which we sample. Proof: $\mathbb{E}(\bar{X}) = \frac{1}{n}\mathbb{E}(X_1) + \frac{1}{n}\mathbb{E}(X_2) + ... + \frac{1}{n}\mathbb{E}(X_n) = \frac{1}{n}\mu + \frac{1}{n}\mu + ... + \frac{1}{n}\mu = \mu$ ■

That's why $\bar{X}$ is often used as an `estimator` of μ, when its value is unknown.

Similarly, $\text{Var}(\bar{X}) = (\frac{1}{n})^2\text{Var}(X_1) + (\frac{1}{n})^2\text{Var}(X_2) + ... + (\frac{1}{n})^2\text{Var}(X_n) = (\frac{1}{n})^2\sigma^2 + (\frac{1}{n})^2\sigma^2 + ... + (\frac{1}{n})^2\sigma^2 = \frac{\sigma^2}{n}$ where σ is the standard deviation of the original distribution. This implies that

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

The standard deviation of $\bar{X}$ (sometimes called the standard error of $\bar{X}$) is $\sqrt{n}$ times smaller than that of the original distribution. Note that the standard error tends to zero as sample size increases.

Now, how about the *shape* of the $\bar{X}$-distribution, how does it relate to the shape of the sampled distribution? The surprising answer is: it doesn't. For n bigger than say 5, the distribution of $\bar{X}$ quickly approaches the *same* regular shape, regardless of how the original distribution looked like.

Now, consider a `random independent sample` of size n from a **bi-variate** distribution, where (X_1, Y_1), (X_2, Y_2), ... (X_n, Y_n) are the individual *pairs* of observations. Then, $\mathrm{Var}(X_1 + X_2 + ... + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + ... + \mathrm{Var}(X_n) \equiv n\,\mathrm{Var}(X)$ and, similarly, $\mathrm{Var}(\sum_{i=1}^{n} Y_i) = n\,\mathrm{Var}(Y)$. Similarly, $\mathrm{Cov}(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i) = \mathrm{Cov}(X_1, Y_1) + \mathrm{Cov}(X_2, Y_2) + ... + \mathrm{Cov}(X_n, Y_n) = n\,\mathrm{Cov}(X, Y)$.

All this implies that the correlation coefficient between $\sum_{i=1}^{n} X_i$ and $\sum_{i=1}^{n} Y_i$ equals $\frac{n\,\mathrm{Cov}(X,Y)}{\sqrt{n\,\mathrm{Var}(X)} \cdot \sqrt{n\,\mathrm{Var}(Y)}} \equiv \rho_{xy}$ (the correlation between a single X and Y pair). The same is true for the corresponding sample means $\bar{X} \equiv \frac{\sum_{i=1}^{n} X_i}{n}$ and $\bar{Y} \equiv \frac{\sum_{i=1}^{n} Y_i}{n}$, why?

Conditional expected value

is, simply put, an expected value computed using the corresponding *conditional* distribution, e.g.

$$\mathbb{E}(X|Y = 1) = \sum_i i \times f_x(i \mid Y = 1)$$

etc.

EXAMPLE: Using our old bivariate distributions

$$\begin{array}{cc|c|c|c|c} & & \multicolumn{3}{c}{X =} & \\ & & 1 & 2 & 3 & \\ \hline Y = & 0 & 0.1 & 0 & 0.3 & 0.4 \\ \hline & 1 & 0.3 & 0.1 & 0.2 & 0.6 \\ \hline & & 0.4 & 0.1 & 0.5 & \end{array}$$

$\mathbb{E}(X|Y = 1)$ is constructed based on the corresponding conditional distribution

$X \vert Y = 1$	1	2	3
Prob:	$\frac{3}{6}$	$\frac{1}{6}$	$\frac{2}{6}$

by the usual process: $1 \cdot \frac{3}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{2}{6} = 1.8\bar{3}$ (note that this is different from $\mathbb{E}(X) = 2.1$ calculated previously). Similarly $\mathbb{E}(X^2|Y = 1) = 1^2 \cdot \frac{3}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{2}{6} = 4.1\bar{6}$. These imply that $\text{Var}(X|Y = 1) = 4.1\bar{6} - 1.8\bar{3}^2 = 0.8056$. Also, $\mathbb{E}(\frac{1}{X}|Y = 1) = \frac{1}{1} \cdot \frac{3}{6} + \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{2}{6} = 0.69\bar{4}$.

When values of a RV are only *integers* (the case of most of our examples so far), we define its **probability generating function** (PGF) by

$$P(z) = \sum_{\text{All } i} z^i \cdot \Pr(X = i)$$

where z is a parameter.

Differentiating k times and substituting $z = 1$ yields

$$\left.\frac{d^k P(z)}{dz^k}\right|_{z=1} = \mathbb{E}[X \cdot (X-1) \cdot \ldots \cdot (X-k+1)]$$

(the so called k^{th} factorial moment). For mutually *independent* RVs, we also get

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

Furthermore, given $P(z)$, we can recover the distribution by Taylor-expanding $P(z)$ - the coefficient of z^i is $\Pr(X = i)$.

EXAMPLES:

- For the distribution of our last example, we get $P(z) = \sum_{i=1}^{\infty} z^i \cdot (\frac{1}{2})^i = \frac{z}{2} \cdot \frac{1}{1-\frac{z}{2}} = \frac{z}{2-z}$ (note that this can be also obtained from the MGF by $e^t \to z$).

$$\begin{aligned} P'(z) &= \left.\frac{2}{(2-z)^2}\right|_{z=1} = 2 \\ P''(z) &= \left.\frac{4}{(2-z)^3}\right|_{z=1} = 4 \end{aligned}$$

giving us the mean of 2 (check) and the variance of $4 + 2 - 2^2 = 2$ (check).

- What is the distribution of the total number of dots when rolling 3 dice? Well, the PGF of X_1, X_2 and X_3 (each) is

$$\frac{z + z^2 + z^3 + z^4 + z^5 + z^6}{6}$$

since they are *independent* RVs, the PGF of their sum is simply

$$\left(\frac{z + z^2 + z^3 + z^4 + z^5 + z^6}{6}\right)^3$$

Expanding (easy in Maple), we get: $\frac{1}{216}z^3 + \frac{1}{72}z^4 + \frac{1}{36}z^5 + \frac{5}{108}z^6 + \frac{5}{72}z^7 + \frac{7}{72}z^8 + \frac{25}{216}z^9 + \frac{1}{8}z^{10} + \frac{1}{8}z^{11} + \frac{25}{216}z^{12} + \frac{7}{72}z^{13} + \frac{5}{72}z^{14} + \frac{5}{108}z^{15} + \frac{1}{36}z^{16} + \frac{1}{72}z^{17} + \frac{1}{216}z^{18}$ ■

Note that a RV can have **infinite** expected value (not too common, but it can happen).

EXAMPLE: Suppose you flip a coin till the first head appears, and you win \$2 if this takes only one flip, \$4 if it takes two flips, \$8 if it takes 3 flips, etc. (the amount always doubles - i.e. you get \$2 for each of the first flips, \$4 for the third, \$8 for the fourth, etc.). What is the expected value of your win?

Clearly, you win $Y = 2^X$ dollars. $\mathbb{E}(Y) = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2^2} + 8 \cdot \frac{1}{2^3} + ... = 1 + 1 + 1 + ... = \infty$.