Basics of Probabilistic Analysis for the APC-Lecture

Emo Welzl, ETH Zürich

Winter Semester 07

I assume that you have some preknowledge in probability theory, but let us recapitulate the small subset of it which we will need for our purposes. We will employ very concrete probability theory (as opposed to abstract probability theory); we simply use it as a tool. When writing this I borrowed much from Chapter 8 in the book *Concrete Mathematics, A Foundation for Computer Science* by Ronald L. Graham, Donald E. Knuth and Oren Patashnik (Addison-Wesley (1989)). A book to be recommended not only for that chapter.

We restrict ourselves to discrete probability spaces, and we will omit 'discrete' from now on. Roughly speaking, such a probability space consists of a (possibly infinite) set of things that can happen, each of which gets assigned a probability that it happens. This mapping is called probability distribution, since it distributes the value 1 among the things that can happen.

Definition 1 (Probability Space) *A* probability space *is a pair* $(\Omega, \Pr)$ *where* Ω *is a set and* $\Pr$ *is a mapping* $\Omega \to \mathbf{R}_0^+$ *such that*

$$\sum_{\omega\in\Omega} \Pr[\omega] = 1 \ .$$

Every subset $\mathcal{E}$ *of* Ω *is called an* event*, and the mapping* $\Pr$ *is extended to events by setting*

$$\Pr[\mathcal{E}] := \sum_{\omega\in\mathcal{E}} \Pr[\omega] \ .$$

The elements in Ω *are the* elementary events*. If* Ω *is finite and* $\Pr[\omega] = \frac{1}{|\Omega|}$ *for all* $\omega \in \Omega$*, then* $\Pr$ *is called* uniform distribution on Ω*. We use* Ω^+ *for the set of elementary events with positive probability,*

$$\Omega^+ := \{\omega \in \Omega \mid \Pr[\omega] > 0\} \ .$$

Rolling dice. Consider the six sides of a die denoted by

$$\mathrm{D} = \{\boxed{1}, \boxed{2}, \boxed{3}, \boxed{4}, \boxed{5}, \boxed{6}\} \ .$$

D models the top side of the die as it lands on the table in an experiment. We consider fair dice, i.e. $\Pr[\mathrm{d}] = \frac{1}{6}$ for all $\mathrm{d} \in \mathrm{D}$. The pair $(\mathrm{D}, \Pr)$ is a probability space with uniform distribution.

For example, $\mathrm{D}_{\text{even}} = \{\boxed{2}, \boxed{4}, \boxed{6}\}$ is the event of having an even number of spots on the top side; $\Pr[\mathrm{D}_{\text{even}}] = 3 \times \frac{1}{6} = \frac{1}{2}$.

Rolling a pair of fair dice is modeled by the set

$$\mathbb{D} := \mathrm{D}^2 = \{\boxed{1}\boxed{1}, \boxed{1}\boxed{2}, \boxed{1}\boxed{3}, \ldots, \boxed{6}\boxed{5}, \boxed{6}\boxed{6}\}$$

of 36 elementary events with the uniform distribution. Note that the two dice are assumed to be distinguishable, say one as the first die, and the other as the second. For the event

$$\mathbb{D}_{=} = \{\boxed{1}\boxed{1}, \boxed{2}\boxed{2}, \boxed{3}\boxed{3}, \boxed{4}\boxed{4}, \boxed{5}\boxed{5}, \boxed{6}\boxed{6}\}$$

we have $\Pr[\mathbb{D}_{=}] = 6 \times \frac{1}{36} = \frac{1}{6}$, and we say that the probability of having the same number of spots on both dice is $\frac{1}{6}$. The event, $\mathbb{D}_{\neq}$, of having a distinct number of spots on the dice is the event complementary to $\Pr[\mathbb{D}_{=}]$; hence, $\Pr\left[\mathbb{D}_{\neq}\right] = 1 - \Pr[\mathbb{D}_{=}] = \frac{5}{6}$. The event $\mathbb{D}_{\neq}$ partitions into the event $\mathbb{D}_{<}$ of having more spots on the second die than on the first, and $\mathbb{D}_{>} = \mathbb{D}_{\neq} \setminus \mathbb{D}_{<}$. $\mathbb{D}_{>}$ and $\mathbb{D}_{<}$ have the same cardinality because of the bijection $\mathrm{d}\mathrm{d}' \mapsto \mathrm{d}'\mathrm{d}$. Hence,

$$\Pr[\mathbb{D}_{>}] = \Pr[\mathbb{D}_{<}] = \frac{\Pr\left[\mathbb{D}_{\neq}\right]}{2} = \frac{5}{12} \ .$$

Flipping coins. Another classical probability space is that of a coin falling on one of its two sides, which results in head or tail with some given probability. Let us use $\mathrm{C} = \{\text{Ⓗ}, \text{Ⓣ}\}$ for the set of elementary events, and let $\Pr[\text{Ⓗ}] = p$ and $\Pr[\text{Ⓣ}] = 1 - p$ for some $p \in \mathbf{R}$, $0 < p < 1$. If $p = \frac{1}{2}$, then we call the coin *fair*; otherwise, it is called *biased*. What if we want to model the experiment of repeatedly flipping a coin until we end up seeing head for the first time? Then

$$\mathrm{C}' = \{\underbrace{\text{Ⓗ}}_{e_0}, \underbrace{\text{ⓉⒽ}}_{e_1}, \underbrace{\text{ⓉⓉⒽ}}_{e_2}, \ldots\} \cup \{\underbrace{\text{ⓉⓉⓉ}\cdots}_{e_\infty}\}$$

where we introduced some convenient shorthands for the elementary events. Here $\Pr[e_i] = p(1-p)^i$, for $i \in \mathbf{N}_0$, and $\Pr[e_\infty] = 0$. At this point accept this as a definition and check that indeed $\sum_{i=0}^{\infty} p(1-p)^i = 1$. Let $\mathrm{C}'_0 = \{e_i \mid i \text{ even}\}$,

the event of waiting an even number of tails until we succeed to see a head. Let $C'_1 = \{e_i \mid i \text{ odd}\}$. We have

$$\Pr[C'_0] = \sum_{i=0}^{\infty} p(1-p)^{2i} = p \sum_{i=0}^{\infty} \left((1-p)^2\right)^i = \frac{p}{1-(1-p)^2} = \frac{1}{2-p} \ .$$

Since C'_1 is – apart from a zero probability elementary event – the event complementary to C'_0, we have

$$\Pr[C'_1] = 1 - \frac{1}{2-p} = \frac{1-p}{2-p} \ .$$

Definition 2 (Random variable) *Given a probability space* $(\Omega, \Pr)$*, a* random variable *is a real-valued function defined on the elementary events of a probability space, i.e.*

$$X : \Omega \to \mathbf{R} \ .$$

If $\mathcal{E}$ *is an event, then*

$$\omega \mapsto [\omega \in \mathcal{E}]$$

is the indicator variable *for event* $\mathcal{E}$*. (The fact that a random variable has to be real-valued is a restriction we apply here. In general, the image of a random variable may be just any set.)*

For the probability space of a single rolling die, we could consider the random variable X^{top} that maps the top side to the number of spots we see on this side

$$X^{\text{top}} : \boxed{1} \mapsto 1, \ \boxed{2} \mapsto 2, \ \boxed{3} \mapsto 3, \ \boxed{4} \mapsto 4, \ \boxed{5} \mapsto 5, \ \boxed{6} \mapsto 6 \ ,$$

or we could map to the number of spots on the invisible side sitting on the table

$$X_{\text{bot}} : \boxed{1} \mapsto 6, \ \boxed{2} \mapsto 5, \ \boxed{3} \mapsto 4, \ \boxed{4} \mapsto 3, \ \boxed{5} \mapsto 2, \ \boxed{6} \mapsto 1 \ .$$

The mapping

$$Y : \boxed{1} \mapsto 0, \ \boxed{2} \mapsto 1, \ \boxed{3} \mapsto 0, \ \boxed{4} \mapsto 1, \ \boxed{5} \mapsto 0, \ \boxed{6} \mapsto 1$$

is the indicator variable for the event of seeing an even number of spots (previously denoted by D_{even}).

Note that X^{top} and X_{bot} depend on each other in the following sense. Suppose we know $X^{\text{top}}(d)$ for some $d \in D$, without knowing d itself, then we know also $X_{\text{bot}}(d)$, because $X^{\text{top}}(d) + X_{\text{bot}}(d) = 7$ for all $d \in D$. We write this as

$$X^{\text{top}} + X_{\text{bot}} = 7$$

for short, omitting the '(d)'. Similarly, X^{top} depends on Y, although not as explicitly as X^{top} depends on X_{bot}. Namely, $Y = 1$ tells us something about X^{top}: $Y(d) = 1 \Rightarrow X^{\text{top}}(d) \in \{2, 4, 6\}$, which we abbreviate as

$$Y = 1 \Rightarrow X^{\text{top}} \in \{2, 4, 6\} .$$

In contrast to this consider two random variables X_1 and X_2 on the space of two rolling dice. X_1 maps to the number of spots on the first die and X_2 to the number of spots on the second die. If $X_1(\omega) = 3$, say, we cannot give a better prediction for the value of $X_2(\omega)$. The same is true for any value possibly attained by X_1. So X_1 and X_2 have completely independent behavior.

For a last example in this context, let Z be the indicator variable for the event that the number of spots on the first die is at most the number of spots on the second die; we could write this as $Z := [X_1 \leq X_2]$. Now, $Z(\omega) = 1$ for some $\omega \in \mathbb{D}$ still allows all possible outcomes of $X_2(\omega)$. However, X_2 depends on Z in the sense that $Z = 1$ makes it more (and most) likely that $X_2 = 6$. That is, knowledge of Z allows a better prediction of X_2 (again, without seeing the underlying event).

Next we will formally capture this intuitive notion of independence.

Definition 3 (Independence) *Let X and Y be random variables defined on a common probability space. We say that X and Y are* independent random variables *if*

$$\Pr[X = x \wedge Y = y] \;=\; \Pr[X = x] \cdot \Pr[Y = y]$$

for all $x, y \in \mathbf{R}$. A collection X_i, $1 \leq i \leq n$ of random variables on a common probability space is called mutually independent *if*

$$\begin{aligned} &\Pr[X_{i_1} = x_{i_1} \wedge X_{i_2} = x_{i_2} \wedge \cdots \wedge X_{i_k} = x_{i_k}] \quad = \\ &\Pr[X_{i_1} = x_{i_1}] \cdot \Pr[X_{i_2} = x_{i_2}] \cdot \cdots \cdot \Pr[X_{i_k} = x_{i_k}] \end{aligned}$$

for all $k \in \{2..n\}$, $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, and $(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) \in \mathbf{R}^k$.

Again, we have used some jargon: '$X = x$' short for the event $\{\omega \in \Omega \mid X(\omega) = x\}$, '$X = x \wedge Y = y$' short for $\{\omega \in \Omega \mid X(\omega) = x \wedge Y(\omega) = y\}$ etc.

It is important to realize that mutual independence is different from pairwise independence. For an example, consider the probability space of a pair of fair coins,

$$C^2 = \{Ⓣ Ⓣ, Ⓣ Ⓗ, Ⓗ Ⓣ, Ⓗ Ⓗ\}$$

with uniform distribution. Now we define three indicator variables

$$\begin{aligned} H_1 &:= [\text{first coin shows } Ⓗ] , \\ H_2 &:= [\text{second coin shows } Ⓗ] , \quad \text{and} \\ H_3 &:= [\text{exactly one coin shows } Ⓗ] . \end{aligned}$$

All three variables attain both 0 and 1 with probability $\frac{1}{2}$. We can verify that the variables are pairwise independent, but

$$\Pr[\text{first coin shows Ⓗ} \wedge \text{second coin shows Ⓗ} \wedge \text{exactly one coin shows Ⓗ}] = 0$$

and not $\frac{1}{8}$ as required for mutual independence.

Definition 4 (Expectation) *Let X be a random real-valued variable. The* expectation *(*expected value*,* mean*) of X is defined as*

$$\mathrm{E}[X] := \sum_{x \in X(\Omega^+)} x \cdot \Pr[X = x] \tag{1}$$

provided this infinite sum exists.

For the example of rolling dice we have

$$\mathrm{E}\left[X^{\mathrm{top}}\right] = \sum_{i=1}^{6} i \frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3.5 \tag{2}$$

or for the random variable X^{top^2}

$$\mathrm{E}\left[X^{\mathrm{top}^2}\right] = \sum_{i=1}^{6} i^2 \frac{1}{6} = \frac{1+4+9+16+25+36}{6} = \frac{91}{6} = 15.16\ldots$$

Note that this was just another shorthand. We used X^{top^2} for the random variable

$$\omega \mapsto (X^{\mathrm{top}}(\omega))^2 \ .$$

Observe that in our example $(\mathrm{E}[X^{\mathrm{top}}])^2 = \frac{49}{4} \neq \frac{91}{6} = \mathrm{E}\left[X^{\mathrm{top}^2}\right]$. Also, if X and Y are random variables, then we cannot expect $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$. Here XY stands for the random variable $\omega \mapsto X(\omega)Y(\omega)$.

Consider X^{top} and X_{bot} as defined for a single die. Then

$$\mathrm{E}\left[X^{\mathrm{top}} X_{\mathrm{bot}}\right] = \frac{1\times 6 + 2\times 5 + 3\times 4 + 4\times 3 + 5\times 2 + 6\times 1}{6} = \frac{28}{3} = 9.33\ldots$$

which is obviously not equal to $\mathrm{E}[X^{\mathrm{top}}]\,\mathrm{E}[X_{\mathrm{bot}}] = \frac{49}{4} = 12.25$.

However, when it comes to linear functions of random variables, we have the following lemma, which is absolutely central for our investigations!

Lemma 1 (Linearity of Expectation) *Let X and Y be random variables defined on a common probability space and let $c \in \mathbf{R}$. Then*

$$\mathrm{E}[cX] = c\,\mathrm{E}[X] \quad \textit{and} \quad \mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] \ ,$$

provided $\mathrm{E}[X]$ and $\mathrm{E}[Y]$ exist.

Here is one typical route along which we will use the linearity of expectation. Recall the experiment of repeatedly flipping a coin until we see head for the first time, assuming that the coin flips are independent and head appears with probability p, $0 < p < 1$. What is the expectation for the number, X, of tails we see? Denote by X_i, $i \in \mathbf{N}$, the indicator variable that we see a tail in the ith round without having seen a head before. We have $X = \sum_{i=1}^{\infty} X_i$. $\Pr[X_i = 1] = (1-p)^i$, since this is a conjunction of i independent trials with success probability $(1-p)$. Hence,

$$\mathrm{E}[X_i] = 1 \times \Pr[X_i = 1] + 0 \times \Pr[X_i = 0] = \Pr[X_i = 1] = (1-p)^i \ ,$$

and[1]

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^{\infty} X_i\right] = \sum_{i=1}^{\infty} \mathrm{E}[X_i] = \sum_{i=1}^{\infty} (1-p)^i = \frac{1}{1-(1-p)} - 1 = \frac{1-p}{p} \ .$$

Note that the X_i's are not independent: $X_j = 1 \Rightarrow X_i = 1$ for all $i < j$, and so $\Pr[X_i = 1 \wedge X_j = 1] = (1-p)^j \neq (1-p)^{i+j}$.

There is a direct way of deriving this expectation, since we know the probabilities $\Pr[X = i]$, $i \in \mathbf{N}_0$. But in many instances we will appreciate that it is possible to determine the expectation without knowing the distribution.

The lemma on the linearity of expectation made no request for independence, while this is required for a similar statement about the product of random variables.

Lemma 2 (Product of Independent Random Variables) *Let X and Y be two* independent *random variables defined on a common probability space. Then*

$$\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y] \ ,$$

provided $\mathrm{E}[X]$ and $\mathrm{E}[Y]$ exist.

Here is a small 'triviality' which is the core of the so-called probabilistic method, where one proves the existence of certain objects by analyzing random objects.

Lemma 3 (Existence from Expectation) *Let X be a random variable on a probability space $(\Omega, \Pr)$ for which the expectation $\mathrm{E}[X]$ exists. Then there exist elementary events ω_1 and ω_2 with*

$$X(\omega_1) \leq \mathrm{E}[X] \quad \textit{and} \quad X(\omega_2) \geq \mathrm{E}[X] \ .$$

Here is another simple fact which we will employ for deriving estimates for the probability that a random variable exceeds a certain value – so-called tail estimates.

[1] Here we have to say "provided all expectations and sums involved exist!"

Lemma 4 (Markov's Inequality) *Let X be a nonnegative random variable (i.e. $X(\Omega) \subseteq \mathbf{R}_0^+$) for which $\mathbb{E}[X] > 0$ exists. Then, for all $\lambda \in \mathbf{R}^+$,*

$$\Pr[X \geq \lambda \mathbb{E}[X]] \leq \frac{1}{\lambda}.$$

Equality holds iff $X(\Omega^+) \subseteq \{0, \lambda \mathbb{E}[X]\}$.

Proof Let $t \in \mathbf{R}^+$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in X(\Omega^+)} x \cdot \Pr[X = x] \\
&= \sum_{x \in X(\Omega^+),\, x<t} \underbrace{x}_{\geq 0} \cdot \Pr[X = x] + \sum_{x \in X(\Omega^+),\, x \geq t} \underbrace{x}_{\geq t} \cdot \Pr[X = x] \\
&\geq t \cdot \sum_{x \in X(\Omega^+),\, x \geq t} \Pr[X = x] \\
&= t \cdot \Pr[X \geq t]
\end{aligned}$$

That is,

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}, \quad \text{for all } t \in \mathbf{R}^+.$$

Moreover, the inequality is strict iff there exists an $x \in X(\Omega^+)$ with $0 < x < t$, or there exists an $x \in X(\Omega^+)$ with $x > t$. It follows that both inequalities are identities iff $x \in X(\Omega^+)$ implies $x \in \{0, t\}$.

Now set $t = \lambda \mathbb{E}[X]$ to conclude the statement of the lemma. □

We close this section with a short discussion of conditional probabilities and expectations.

Suppose somebody, call him Mr. McChance, offers you the following deal. First, you get 3.5 swiss francs. Then you have to roll two dice. If the second die shows a larger number of spots than the first one, you have to return that number (of spots on the second) of francs to friendly Mr. McChance; otherwise you have to return 2.5 francs to him. That may look quite attractive, at first glance, since the expected number of spots on the top face of the second rolling die is 3.5. And we even have some chance of paying 2.5 only. But then you play the game several times, and it looks like you are losing. You are getting worried, and decide upon a thorough investigation of the game.

In order to analyze our expected gain or loss in the game, we have to distinguish two cases: The event $\mathbb{D}_{\geq}$ of the first die showing at least as many spots as the second, and the complementary event

$$\begin{aligned}\mathbb{D}_< \;= \{\;& \boxed{1}\boxed{2}, \boxed{1}\boxed{3}, \boxed{1}\boxed{4}, \boxed{1}\boxed{5}, \boxed{1}\boxed{6},\\ & \boxed{2}\boxed{3}, \boxed{2}\boxed{4}, \boxed{2}\boxed{5}, \boxed{2}\boxed{6},\\ & \boxed{3}\boxed{4}, \boxed{3}\boxed{5}, \boxed{3}\boxed{6},\\ & \boxed{4}\boxed{5}, \boxed{4}\boxed{6},\\ & \boxed{5}\boxed{6}\;\}\end{aligned}$$

Here, of course, we see the pitfall of the procedure. *Given the event, that we have to pay the number of spots on the second die*, this number tends to be large – there is no configuration for that number to be 1, one for it to be 2, ..., while there are 5 for it to be 6. Within the space of the 15 possible elementary events in $\mathbb{D}_<$, assuming uniform distribution among them, we expect to pay

$$2 \times \frac{1}{15} + 3 \times \frac{2}{15} + 4 \times \frac{3}{15} + 5 \times \frac{4}{15} + 6 \times \frac{5}{15} = \frac{70}{15} = 4.66\ldots \;.$$

So much for the bad news. But we may be lucky, $\mathbb{D}_\geq$ occurs (the chance for this to happen is 21 in 36 cases), and we have to pay 2.5 francs. We weight the two cases according to their probabilities to occur, and conclude that the expected number of francs we have to pay back is

$$\frac{70}{15} \times \frac{15}{36} + \frac{5}{2} \times \frac{21}{36} = \frac{245}{72} = 3.402\ldots \;.$$

So, after all, we have an expected gain of roughly 0.1 Swiss francs in the game. We can conclude that either (i) we made a mistake in our calculation, (ii) Mr. McChance brought loaded dice with him, (iii) bad luck, (iv) etc.

The analysis we just performed employs conditional probabilities – an essential tool in the analysis of randomized algorithms.

Definition 5 (Conditional Probabilities) *Let* A *and* B *be events in a probability space with* $\Pr[B] > 0$. *The* conditional probability of A, given B, *is defined to be*

$$\Pr[A \mid B] := \frac{\Pr[A \cap B]}{\Pr[B]} \ .$$

(In particular, let X *and* Y *be random variables defined on a common probability space. Then the conditional probability of the event* $X = x$, *given the event* $Y = y$, *is*

$$\Pr[X = x \mid Y = y] = \frac{\Pr[X = x \wedge Y = y]}{\Pr[Y = y]}$$

for all $x, y \in \mathbf{R}$ *with* $\Pr[Y = y] > 0$.*)*
Let X *be a random variable and* B *be an event in a common probability space,* $\Pr[B] > 0$. *Then* $X \mid B$ *is the random variable obtained as the restriction of* X *to the probability space* $(B, \Pr')$ *with*

$$\Pr' : \quad \omega \mapsto \frac{\Pr[\omega]}{\Pr[B]} \ .$$

Conditional probabilities usually create some 'notational confusion'. Let's just add to this by asking to verify the identity

$$\Pr[X = x \mid Y = y] = \Pr'((X \mid Y = y) = x) \ .$$

Anyway, a random variable has an expectation, and so has $X \mid Y = y$.

$$\begin{aligned} \mathrm{E}[X \mid Y = y] &= \sum_{x \in (X \mid Y = y)((Y = y)^+)} x \Pr[(X \mid Y = y) = x] \\ &= \sum_{x \in X(\Omega^+)} x \Pr[X = x \mid Y = y] \ . \end{aligned}$$

provided the sum exists.

In the analysis of the game with McChance, we analyzed the random variable X for the amount we have to pay back after the experiment. Let Y be the indicator variable for the event $\mathbb{D}_{\geq}$. Then $\mathrm{E}[X \mid Y = 1] = 2.5$ and we calculated $\mathrm{E}[X \mid Y = 0] = 4.66\ldots$. The justification for our final step in the derivation of $\mathrm{E}[X]$ is given in the lemma below.

Lemma 5 (Laws of Total Probability and Expectation) *Let* X *and* Y *be two random variables on a common probability space* $(\Omega, \Pr)$*. Then*

$$\Pr[X = x] = \sum_{y \in Y(\Omega^+)} \Pr[X = x \mid Y = y] \Pr[Y = y] \tag{3}$$

for $x \in \mathbf{R}$*, and*

$$\mathrm{E}[X] = \sum_{y \in Y(\Omega^+)} \mathrm{E}[X \mid Y = y] \Pr[Y = y] \ , \tag{4}$$

provided $\mathrm{E}[X]$ *exists.*

Proof

$$\begin{aligned}
&\sum_{y \in Y(\Omega^+)} \Pr[X = x \mid Y = y] \Pr[Y = y] \\
&= \sum_{y \in Y(\Omega^+)} \frac{\Pr[X = x \wedge Y = y]}{\Pr[Y = y]} \Pr[Y = y] \\
&= \sum_{y \in Y(\Omega^+)} \Pr[X = x \wedge Y = y] \\
&= \Pr[X = x]
\end{aligned}$$

since every elementary event $\omega \in \Omega^+$ with $X(\omega) = x$ is mapped by Y to a unique $y \in Y(\Omega^+)$.

$$\begin{aligned}
&\sum_{y \in Y(\Omega^+)} \mathrm{E}[X \mid Y = y] \Pr[Y = y] \\
&= \sum_{y \in Y(\Omega^+)} \left(\sum_{x \in X(\Omega^+)} x \Pr[X = x \mid Y = y] \right) \Pr[Y = y] \\
&= \sum_{x \in X(\Omega^+)} x \sum_{y \in Y(\Omega^+)} \Pr[X = x \mid Y = y] \Pr[Y = y] \\
&= \sum_{x \in X(\Omega^+)} x \Pr[X = x] \\
&= \mathrm{E}[X]
\end{aligned}$$

□

In the game with McChance, we derived

$$\mathrm{E}[X] = \underbrace{\mathrm{E}[X \mid Y = 0]}_{70/15} \underbrace{\Pr[Y = 0]}_{15/36} + \underbrace{\mathrm{E}[X \mid Y = 1]}_{5/2} \underbrace{\Pr[Y = 1]}_{21/36} \ .$$