A Note on “Reordering” Infinite Sums

Lasse Kliemann
`lki@informatik.uni-kiel.de`

Permanent ID of this document:
`7898f04f-8992-455a-bc73-e50b7ab79e75`

Revision 5753

Abstract

In discrete probability theory, we often have to deal with infinite sums. It usually is requested that such a sum converges unconditionally, i.e., in any ordering and always to the same limit. (This is automatically the case if the sum converges in one ordering and has only got non-negative addends.) Unconditional convergence allows us to reorder infinite sums as we like. However, not everything that may look like a reordering at first glance is one in fact. A common move is to *partition* the addends and to take the limit over each class of the partition, then taking the limit over all those limits. In this note, we prove a lemma stating that this leads to the same limit as the original sum. In text books, this question often is not addressed but the partitioning is incorrectly described as a “reordering”.

For example, we encounter such a situation when we prove the well-known identity $\mathbb{E}(X) = \sum_{x \in \mathbb{R}} x\, \mathbb{P}(X = x)$ for a real-valued random variable X, expectation being defined as $\mathbb{E}(X) \coloneqq \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$. In addition to proving the lemma, we will in this note elaborate on this example and give two other examples from elementary discrete probability theory where the lemma is of use.

Unconditional convergence allows us to reorder infinite sums as we like. Formally, a reordering of a sum $\sum_{k=1}^{\infty} a_k$ is a sum $\sum_{k=1}^{\infty} a_{\alpha(k)}$, where $\alpha : \mathbb{N} \longrightarrow \mathbb{N}$ is a bijective mapping. Not everything that may look like a reordering at first glance is one in fact. A common move is to partition the addends and to take the limit over each class of the partition, then taking the limit over all those limits. For example, this is utilized to prove the well-known identity $\mathbb{E}(X) = \sum_{x \in \mathbb{R}} x\, \mathbb{P}(X = x)$ for a real-valued random variable X. The expectation is defined as $\mathbb{E}(X) \coloneqq \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$, where $X : \Omega \longrightarrow \mathbb{R}$ and the sum is requested to converge unconditionally. We write:

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) \overset{(*)}{=} \sum_{x \in \mathbb{R}} \sum_{\substack{\omega \in \Omega: \\ X(\omega) = x}} X(\omega)\mathbb{P}(\{\omega\})$$

$$= \sum_{x \in \mathbb{R}} \sum_{\substack{\omega\in\Omega:\\ X(\omega)=x}} x\,\mathbb{P}(\{\omega\}) = \sum_{x\in\mathbb{R}} x \sum_{\substack{\omega\in\Omega:\\ X(\omega)=x}} \mathbb{P}(\{\omega\}) = \sum_{x\in\mathbb{R}} x\,\mathbb{P}(X=x) \enspace .$$

At $(*)$ we do partitioning. Instead of taking the limit over all $\omega \in \Omega$, we take, for each $x \in \mathbb{R}$, the limit over all $\omega \in \Omega$ that map to x under X, and finally take the limit over all those limits (indexed by $x \in \mathbb{R}$). This partitioning is most often described as "reordering" in text books and not explained any further, most probably because reordering is covered by unconditional convergence. But unless $X^{-1}(x) = \{\omega \in \Omega;\ X(\omega) = x\}$ has finite cardinality for each $x \in \mathbb{R}$, it is not just a reordering. In this note, we prove a general lemma stating that this step is correct nonetheless.

Before we turn to the lemma and the proof, we give two more examples of applications in discrete probability theory. Let Ω be countable and $p : \Omega \longrightarrow [0,1]$ a stochastic vector, i.e., $\sum_{\omega\in\Omega} p(\omega) = 1$. Then $\mathbb{P} : 2^{\Omega} \longrightarrow [0,1], A \mapsto \sum_{\omega\in A} p(\omega)$ is a probability measure. While proving this, we have to consider a series of pairwise disjoints events $(A_n)_{n\in\mathbb{N}}$ and show $\mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$. This is again a partitioning, not a reordering. We have to transform $\sum_{\omega\in A_1\cup A_2\cup\ldots} p(\omega)$ into $\sum_{\omega\in A_1} p(\omega) + \sum_{\omega\in A_2} p(\omega) + \ldots$. That is, we have to show that taking the limit over the union $\bigcup_{n\in\mathbb{N}} A_n$ is the same as taking the limit over each set A_n, $n \in \mathbb{N}$, and finally taking the limit over all those limits.

As a third example, we consider the product measure. Let $(\Omega_1, 2^{\Omega_1}, \mathbb{P}_1)$ and $(\Omega_2, 2^{\Omega_2}, \mathbb{P}_2)$ be discrete proability spaces, and denote $\Omega \coloneqq \Omega_1 \times \Omega_2$. Then $p : \Omega \longrightarrow [0,1], (\omega_1,\omega_2) \mapsto \mathbb{P}(\{\omega_1\})\,\mathbb{P}(\{\omega_2\})$ is a stochastic vector. The measure defined by p is called the product measure of $\mathbb{P}_1$ and $\mathbb{P}_2$. Proving that p is in fact a stochastic vector, we require sum partitioning again. We write:

$$\begin{aligned}
\sum_{\omega\in\Omega} p(\omega) &= \sum_{(\omega_1,\omega_2)\in\Omega} p(\omega_1,\omega_2) = \sum_{(\omega_1,\omega_2)\in\Omega} \mathbb{P}_1(\{\omega_1\})\,\mathbb{P}_2(\{\omega_2\}) \\
&\overset{(*)}{=} \sum_{\omega_1\in\Omega_1} \sum_{\omega_2\Omega_2} \mathbb{P}_1(\{\omega_1\})\,\mathbb{P}_2(\{\omega_2\}) = \sum_{\omega_1\in\Omega_1} \mathbb{P}_1(\{\omega_1\}) \sum_{\omega_2\Omega_2} \mathbb{P}_2(\{\omega_2\}) \\
&= \sum_{\omega_1\in\Omega_1} \mathbb{P}_1(\{\omega_1\}) \cdot 1 = 1 \enspace .
\end{aligned}$$

The innocent-looking transformation marked with $(*)$ is the partitioning. Instead of taking the limit over all $(\omega_1,\omega_2) \in \Omega$, we take, for each $\omega_1 \in \Omega_1$, the limit over all $\omega_2 \in \Omega_2$, and then take the limit over all those limits (indexed by $\omega_1 \in \Omega_1$). Pulling "$\mathbb{P}_1(\{\omega_1\})$" out of the inner sum on the other hand is only an application of the following simple and well-known rule: If a sequence $(a_k)_{k\in\mathbb{N}}$ converges to limit a, then for each $c \in \mathbb{R}$ the sequence $(ca_k)_{k\in\mathbb{N}}$ converges to limit ca.

In case of this last example, we could have instead invoked a classical result, namely the Cauchy product. It yields directly

$$\sum_{(\omega_1,\omega_2)\in\Omega} \mathbb{P}_1(\{\omega_1\})\,\mathbb{P}_2(\{\omega_2\}) = \Big(\sum_{\omega_1\in\Omega_1} \mathbb{P}_1(\{\omega_1\})\Big)\Big(\sum_{\omega_2\in\Omega_2} \mathbb{P}_2(\{\omega_2\})\Big)\ .$$

But there appears to be no way that the Cauchy product could help us in the other two examples.

Now we turn to our lemma and its proof. Our proof consists mainly of technically involved applications of Cauchy's condition. Cauchy's condition states that a series $\sum_{k=1}^{\infty} a_k$ converges if and only if for each $\varepsilon > 0$ there exist n_0 such that

$$\left|\sum_{k=n_1+1}^{n_2} a_k\right| < \varepsilon \quad \text{for all } n_0 \le n_1 \le n_2.$$

We also recall that a sum $\sum_{k=1}^{\infty} a_k$ is called *absolutely convergent* if $\sum_{k=1}^{n} |a_k|$ converges. It is well known that an absolutely convergent sum is unconditionally convergent, and vice versa. We will usually use the term "absolute convergence" instead of "unconditional convergence" from now on.

Lemma. *Let $(a_k)_{k\in\mathbb{N}}$ be a real sequence. Let $(P_i)_{i\in\mathbb{N}}$ be a partition of $\mathbb{N}$, i.e., $\biguplus_{i\in\mathbb{N}} P_i = \mathbb{N}$. For each $i \in \mathbb{N}$ denote $A_i := \{a_k;\ k \in P_i\}$ and assume*

$$A_i \subseteq (-\infty, 0] \quad \textit{or} \quad A_i \subseteq [0,\infty)\ . \tag{1}$$

Then the following two conditions are equivalent:

(a) The series $\sum_{k\in\mathbb{N}} a_k$ is absolutely convergent.

(b) For each $i \in \mathbb{N}$, the series $\sum_{k\in P_i} a_k$ is absolutely convergent; denote s_i its limit. The series $\sum_{i\in\mathbb{N}} s_i$ is absolutely convergent.

If one condition holds – and so both – then $\sum_{k\in\mathbb{N}} a_k = \sum_{i\in\mathbb{N}} s_i$.

Which case in condition (1) holds, is allowed to be different for different i. Note once again that the two series $\sum_{k\in\mathbb{N}} a_k$ and $\sum_{i\in\mathbb{N}} s_i = \sum_{i\in\mathbb{N}}\sum_{k\in P_i} a_k$ are not just reorderings of each other (unless each P_i has finite cardinality).

We show in detail how to apply the lemma to prove

$$\mathbb{E}(X) = \sum_{x\in\mathbb{R}} x\,\mathbb{P}(X = x)\ . \tag{2}$$

We enumerate all samples $\Omega = \{\omega_1, \omega_2, \ldots\}$, then define $a_k := X(\omega_k)\mathbb{P}(\{\omega_k\})$ for each $k \in \mathbb{N}$. This establishes (a) in the lemma. Then we enumerate the image $X(\Omega) = \{x_1, x_2, \ldots\}$. For each $i \in \mathbb{N}$ denote $P_i := \{k \in \mathbb{N};\ X(\omega_k) = x_i\}$. Then $A_i = \{X(\omega_k)\mathbb{P}(\{\omega_k\});\ k \in P_i\} = \{x_i\mathbb{P}(\{\omega_k\});\ k \in P_i\}$. This shows (1) since probabilities are non-negative; whether we have

$A_i \subseteq (-\infty, 0]$ or $A_i \subseteq [0, \infty)$ depends on x_i. Now the lemma states that $\sum_{i \in \mathbb{N}} \sum_{k \in P_i} a_k$ converges to $\mathbb{E}(X)$. We so have

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i \in \mathbb{N}} \sum_{k \in P_i} a_k = \sum_{i \in \mathbb{N}} \sum_{k \in P_i} X(\omega_k) \mathbb{P}(\{\omega_k\}) = \sum_{i \in \mathbb{N}} \sum_{k \in P_i} x_i \, \mathbb{P}(\{\omega_k\}) \\ &= \sum_{i \in \mathbb{N}} x_i \sum_{k \in P_i} \mathbb{P}(\{\omega_k\}) = \sum_{i \in \mathbb{N}} x_i \, \mathbb{P}(X = x_i) = \sum_{x \in \mathbb{R}} x \, \mathbb{P}(X = x) \ . \end{aligned}$$

Proof of the lemma. For each $i \in \mathbb{N}$ we enumerate the indices in P_i by $k_1^{(i)} < k_2^{(i)} < \ldots$. Denote $a_j^{(i)} := a_{k_j^{(i)}}$ for each $j \in \{1, \ldots, |P|\}$ if P_i has finite cardinality and for each $j \in \mathbb{N}$ otherwise. If P_i is finite we set $a_j^{(i)} := 0$ for all $j > |P|$. Then $s_i = \sum_{j \in \mathbb{N}} a_j^{(i)}$ and also $|s_i| = \sum_{j \in \mathbb{N}} |a_j^{(i)}|$ due to (1).

"(a) $\implies$ (b)" Absolute convergence of $\sum_{k \in P_i} a_k$ is clear for each $i \in \mathbb{N}$, since $\sum_{k \in \mathbb{N}} a_k$ is absolutely convergent. (Realize that the sequence $(\sum_{k=1}^{n} |a_k|)_{n \in \mathbb{N}}$ is non-decreasing and bounded.)

Let $\sum_{k \in \mathbb{N}} a_k$ be absolutely convergent to limit a. Then $\sum_{d=2}^{\infty} \sum_{(i,j)\colon i+j=d} a_j^{(i)}$ converges to a, since it is a reordering of that series (the ordering in the inner sum is irrelevant, since it is finite). We use Cauchy's condition to show that $\sum_{i \in \mathbb{N}} |s_i|$ is convergent. Let $\varepsilon > 0$. Then, by Cauchy's condition, there exists d_0 such that

$$\left| \sum_{d=d_1+1}^{d_2} \sum_{\substack{(i,j): \\ i+j=d}} |a_j^{(i)}| \right| < \varepsilon/2 \quad \text{for all } d_0 \le d_1 \le d_2. \tag{3}$$

In particular, we have

$$\left| \sum_{i=i_1+1}^{i_2} \sum_{j=1}^{n} |a_j^{(i)}| \right| < \varepsilon/2 \quad \text{for all } d_0 - 1 \le i_1 \le i_2 \text{ and all } n. \tag{4}$$

To see this, note that "$i+j$" is always large enough in the inner sum, namely at least $d_0 + 1$.

We show that Cauchy's condition also holds for the series in question. Let $d_0 - 1 \le i_1 \le i_2$. We choose n_0 large enough so that

$$\left| |s_i| - \sum_{j=1}^{n} |a_j^{(i)}| \right| < \frac{\varepsilon}{2(i_2 - i_1)} \quad \text{for all } n_0 \le n \text{ and } i \in \{i_1 + 1, \ldots, i_2\}. \tag{5}$$

Then, using any $n \ge n_0$,

$$\sum_{i=i_1+1}^{i_2} |s_i| = \sum_{i=i_1+1}^{i_2} \left(|s_i| - \sum_{j=1}^{n} |a_j^{(i)}| + \sum_{j=1}^{n} |a_j^{(i)}| \right)$$

$$
\begin{aligned}
&\le \left| \sum_{i=i_1+1}^{i_2} \left(|s_i| - \sum_{j=1}^{n} |a_j^{(i)}| \right) \right| + \left| \sum_{i=i_1+1}^{i_2} \sum_{j=1}^{n} |a_j^{(i)}| \right| \\
&\le \sum_{i=i_1+1}^{i_2} \left| |s_i| - \sum_{j=1}^{n} |a_j^{(i)}| \right| + \sum_{i=i_1+1}^{i_2} \sum_{j=1}^{n} |a_j^{(i)}| \\
&\overset{(5)}{<} \frac{\varepsilon}{2} + \sum_{i=i_1+1}^{i_2} \sum_{j=1}^{n} |a_j^{(i)}| \overset{(4)}{<} \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \ .
\end{aligned}
$$

By Cauchy's condition, the series $\sum_{i\in\mathbb{N}} |s_i|$ is convergent.

"(b) $\implies$ (a)" Let $\sum_{i\in\mathbb{N}} s_i$ be absolutely convergent with limit s. We show that the series over $(\sum_{(i,j):\, i+j=d} a_j^{(i)})_{d=2,\dots,\infty}$ converges to s as well. Then we invoke this result on $\sum_{i\in\mathbb{N}} |s_i|$ instead of $\sum_{i\in\mathbb{N}} s_i$. Since then all summands are non-negative, this yields something slightly stronger, namely convergence of $\sum_{d=2}^{\infty} \sum_{(i,j):\, i+j=d} |a_j^{(i)}|$; this is a series over $(|a_k|)_{k\in\mathbb{N}}$ in some order. This implies absolute convergence of $\sum_{k\in\mathbb{N}} a_k$. We also know then that the limit of $\sum_{k\in\mathbb{N}} a_k$ and that of $\sum_{i\in\mathbb{N}} s_i$ are both s.

So all left to show is $\sum_{d=1}^{\infty} \sum_{(i,j):\, i+j=d} a_j^{(i)} = s$. To this end, let $\varepsilon > 0$. First we exploit (absolute) convergence of $\sum_{i\in\mathbb{N}} s_i$. We choose m_0 so that

$$
\left| s - \sum_{i=1}^{m} s_i \right| < \frac{\varepsilon}{3} \quad \text{for all } m_0 \le m \tag{6}
$$

and, using Cauchy's condition and absolute convergence,

$$
\sum_{i=m_1+1}^{m_2} |s_i| < \frac{\varepsilon}{3} \quad \text{for all } m_0 \le m_1 \le m_2. \tag{7}
$$

By (1), we have

$$
\sum_{j\in J} |a_j^{(i)}| \le |s_i| \quad \text{for each } i \text{ and any index set } J. \tag{8}
$$

Having m_0 fixed, we exploit that for each $i \in \{1, \dots, m_0\}$ the series $\sum_{j=1}^{\infty} a_j^{(i)}$ converges to s_i. We choose n_0 large enough so that

$$
\left| s_i - \sum_{j=1}^{n} a_j^{(i)} \right| < \frac{\varepsilon}{3m_0} \quad \text{for all } n_0 \le n \text{ and } i \in \{1, \dots, m_0\}. \tag{9}
$$

Let $d_0 \coloneqq m_0 + n_0$ and $d_0 \le d'$. For later, we note

$$
\begin{aligned}
&\{(d,i);\ 2 \le d \le d' \ \wedge\ m_0 + 1 \le i \le d-1\} \\
=&\{(d,i);\ i+1 \le d \le d' \ \wedge\ m_0 + 1 \le i \le d'-1\}\ .
\end{aligned} \tag{10}
$$

Then:

$$
\begin{aligned}
&\left|s-\sum_{d=2}^{d'}\sum_{\substack{(i,j):\\ i+j=d}} a_j^{(i)}\right| \\
&=\left|s-\sum_{d=2}^{d'}\Big(\sum_{\substack{(i,j):\\ i+j=d\\ i\le m_0}} a_j^{(i)}+\sum_{\substack{(i,j):\\ i+j=d\\ i>m_0}} a_j^{(i)}\Big)\right| \\
&=\left|s-\sum_{i=1}^{m_0}s_i+\left(\sum_{i=1}^{m_0}s_i-\sum_{d=2}^{d'}\sum_{\substack{(i,j):\\ i+j=d\\ i\le m_0}} a_j^{(i)}\right)-\sum_{d=2}^{d'}\sum_{\substack{(i,j):\\ i+j=d\\ i>m_0}} a_j^{(i)}\right| \\
&\overset{(6)}{<}\frac{\varepsilon}{3}+\left|\sum_{i=1}^{m_0}s_i-\sum_{d=2}^{d'}\sum_{\substack{(i,j):\\ i+j=d\\ i\le m_0}} a_j^{(i)}\right|+\left|\sum_{d=2}^{d'}\sum_{\substack{(i,j):\\ i+j=d\\ i>m_0}} a_j^{(i)}\right| \\
&=\frac{\varepsilon}{3}+\left|\sum_{i=1}^{m_0}s_i-\sum_{i=1}^{m_0}\sum_{d=2}^{d'}\sum_{\substack{j:\\ i+j=d}} a_j^{(i)}\right|+\left|\sum_{d=2}^{d'}\sum_{i=m_0+1}^{d-1}\sum_{\substack{j:\\ i+j=d}} a_j^{(i)}\right| \\
&=\frac{\varepsilon}{3}+\left|\sum_{i=1}^{m_0}s_i-\sum_{i=1}^{m_0}\sum_{d=i+1}^{d'} a_{d-i}^{(i)}\right|+\left|\sum_{d=2}^{d'}\sum_{i=m_0+1}^{d-1} a_{d-i}^{(i)}\right| \\
&=\frac{\varepsilon}{3}+\left|\sum_{i=1}^{m_0}s_i-\sum_{i=1}^{m_0}\sum_{j=1}^{d'-i} a_j^{(i)}\right|+\left|\sum_{i=m_0+1}^{d'-1}\sum_{d=i+1}^{d'} a_{d-i}^{(i)}\right| \qquad \text{(10) for 2nd sum} \\
&\le\frac{\varepsilon}{3}+\sum_{i=1}^{m_0}\left|s_i-\sum_{j=1}^{d'-i} a_j^{(i)}\right|+\left|\sum_{i=m_0+1}^{d'-1}\sum_{d=i+1}^{d'} a_{d-i}^{(i)}\right| \\
&\overset{(9)}{\le}\frac{\varepsilon}{3}+\sum_{i=1}^{m_0}\frac{\varepsilon}{3m_0}+\left|\sum_{i=m_0+1}^{d'-1}\sum_{d=i+1}^{d'} a_{d-i}^{(i)}\right| \qquad \begin{array}{l}\text{note that } d'-i\ge\\ (m_0+n_0)-m_0=n_0\\ \text{for } i\le m_0\end{array} \\
&\overset{(8)}{\le}\frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\sum_{i=m_0+1}^{d'-1}|s_i| \\
&\overset{(7)}{\le}\frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3}=\varepsilon\ . \qquad\qquad \square
\end{aligned}
$$