Chuck Holmes (holmes@wustl.edu)

From Vinck, et al, 2012, the equation for PPC estimator P2 is:

$$\widehat{P_2} = \frac{1}{|M|(|M|-1)} \sum_{m \in M} \sum_{l \in L} \left(\frac{\sum_{k=1}^{N_m} \sum_{j=1}^{N_l} U_{k,m} \cdot U_{j,l}}{N_m N_l} \right)$$

In the equation, M is the set of all trials and L is the set of all trials minus trial m.

We can distribute the normalization factors and use the distributive property to make the equation:

$$\widehat{P_2} = \frac{1}{|M|(|M|-1)} \sum_{m \in M} \sum_{l \in L} \left(\left(\frac{1}{N_m} \sum_{k=1}^{N_m} U_{k,m} \right) \cdot \left(\frac{1}{N_l} \sum_{j=1}^{N_l} U_{j,l} \right) \right)$$

We can see these are both averages, so let's simplify:

$$\bar{U}_m = \left(\frac{1}{N_m} \sum_{k=1}^{N_m} U_{k,m} \right)$$

$$\widehat{P_2} = \frac{1}{|M|(|M|-1)} \sum_{m \in M} \sum_{l \in L} (\bar{U}_m \cdot \bar{U}_l)$$

Next, distribute again…

$$\widehat{P_2} = \frac{1}{|M|(|M|-1)} \sum_{m \in M} \left(\bar{U}_m \cdot \left(\sum_{l \in L} \bar{U}_l \right) \right)$$

This would be simpler if we were working with one set and no subsets, so let's add and subtract $\bar{U}_m$:

$$\widehat{P_2} = \frac{1}{|M|(|M|-1)} \sum_{m \in M} \left(\bar{U}_m \cdot \left(\bar{U}_m - \bar{U}_m + \sum_{l \in L} \bar{U}_l \right) \right)$$

$$= \frac{1}{|M|(|M|-1)} \sum_{m \in M} \left(\bar{U}_m \cdot \left(-\bar{U}_m + \sum_{l \in M} \bar{U}_l \right) \right)$$

And then distribute:

$$\widehat{P_2} = \frac{1}{|M|(|M|-1)} \sum_{m \in M} \left(\bar{U}_m \cdot \sum_{l \in M} \bar{U}_l - \bar{U}_m \cdot \bar{U}_m \right)$$

$$= \frac{1}{|M|(|M|-1)} \left(\sum_{m \in M} \left(\bar{U}_m \cdot \sum_{l \in M} \bar{U}_l \right) - \sum_{m \in M} (\bar{U}_m \cdot \bar{U}_m) \right)$$

$$= \frac{1}{|M|(|M|-1)}\left(\left(\sum_{m\in M}\bar{U}_m\right)\cdot\left(\sum_{l\in M}\bar{U}_l\right)-\sum_{m\in M}(\bar{U}_m\cdot\bar{U}_m)\right)$$

And finally, we see that the two sums in the "middle" are identical, so lets change that:

$$= \frac{1}{|M|(|M|-1)}\left(\left(\sum_{m\in M}\bar{U}_m\right)\cdot\left(\sum_{m\in M}\bar{U}_l\right)-\sum_{m\in M}(\bar{U}_m\cdot\bar{U}_m)\right)$$

Finally, let's change those auto-dot-products to magnitude squared operations:

$$= \frac{1}{|M|(|M|-1)}\left(\left|\sum_{m\in M}\bar{U}_m\right|^2-\sum_{m\in M}|\bar{U}_m|^2\right)$$