**THE SQUARE ROOT THEOREM OF MEMORY SHARING**

Peter J. Denning

6/28/20 DRAFT

Early in 2020, someone notified me that he had found a theorem I had proved in my 1968 PhD thesis to be extremely useful in a modern cache memory setting. The theorem said that if a memory is equipartitioned *n* ways (fixed partition multiprogramming), the total memory requirement is $\sqrt{n}$ larger than if the same memory is dynamically shared (working set governed mulitprogramming). This was not a major result in my thesis and I had nearly forgotten it over the years. But given that someone else was interested, I wondered if it might be a useful addition to the "working set analytics" paper I was working on. I went back and reviewed the theorem. I discovered that the theorem rests on an unjustifiable assumption and is weaker than I suggested in my thesis.

The derivation assumed that the amount of memory *y* requested by a job followed a probability density function *f(y)* with mean 0. I assumed (wrongly) that the result was linearly invariant and would work for distributions with mean *m>0*. As I traced the derivation, I saw that the assumption *m=0* was necessary for the conclusion. If *m>0*, the conclusion is weakened. My purpose here is to outline a correct theorem.

The objective is to compare the memory required with equipartition and with optimal sharing, where the memory sizes were chosen to make the probability of a request for memory being denied the same in both cases. Notation:

*y* is the random variable of the request of a single job for memory

*m* is the mean of *y* and $var(y) = \overline{y^2} - m^2 \triangleq s^2$

*s* is the standard deviation of *y*

*n* is number of jobs

*Y* is the random variable of the total request of *n* jobs, $Y = \sum_1^n y_i$

Note that $var(Y) = n\ var(y)$ by statistical independence, $mean(Y) = nm$

*R* is the memory size allocated to one job in the equipartition

*nR* is the total memory allocated for equipartition

*R'* is the total memory size available in shared multiprogramming

**A Basic Inequality**

$$s^2 = \int_0^\infty (u^2 - m^2) f(u) du \geq \int_R^\infty u^2 f(u) du - m^2 \geq R^2 P(y > R) - m^2$$

Thus an upper bound on the exceedance probability *P(y>R)* is

$$P(y > R) \leq \frac{s^2 + m^2}{R^2}$$

This is the probability that a request of size *R* will be denied in the equipartition. For the shared memory, the inequality becomes

$$P(Y > R') \leq \frac{ns^2 + (nm)^2}{R'^2}$$

**Application to the Memory**

To compare the equipartition with shared memory, we want to choose *R* and *R'* so that the exceedance probabilities are the same. For those memory sizes a job experiences the same chance of denial of request under the two memory allocation schemes. Because we do not have the exact expressions for *P(y>R)* and *P(Y>R')*, we approximate by setting their upper bounds equal, giving:

$$\frac{R'^2}{R^2} = \frac{ns^2 + (nm)^2}{s^2 + m^2} = A$$

Thus at the balance point

$$R = \frac{R'}{\sqrt{A}}$$

The ratio of total equipartition memory to dynamic shared memory is

$$\frac{nR}{R'} = \frac{n}{\sqrt{A}}$$

In the case covered in my PhD thesis, *m=0*. That puts *A=n* and the ratio at $\sqrt{n}$, as proved there.

However, in general *m≠0* and *A≠n*. For example, in the totally deterministic case where *s=0*, $A=n^2$ and the ratio is 1. There is no advantage to dynamically shared memory when all jobs require exactly the same amount of memory.

We suspect that for nonzero *s* and *m*, $\sqrt{n}$ is the most optimistic for the ratio and 1 the most pessimistic. Consider these inequalities:

$$n < n\,\frac{s^2 + nm^2}{s^2 + m^2} = A = n^2\,\frac{\frac{s^2}{n} + m^2}{s^2 + m^2} < n^2$$

And so

$$n < A < n^2$$

as claimed.