The most important reason to care about the mean value comes from its connection to estimation by sampling. If we want to measure a population, we first determine a random process for selecting people. This process makes the person's age, income, and so on a random variable whose *mean* equals the *actual average* age or income of the population. So, we can select a random sample of people and calculate the average of people in the sample to estimate the true average in the whole population.

1. Markov's Theorem

gives a generally coarse estimate of the probability that a random value takes a value *much larger* than its mean. The idea can be explained by considering the idea behind the *intelligence quotient* (IQ). IQ is defined to have an average measurement of 100. This implies that at most 1/3 of the population can have an IQ of 300 or more, because if more than a third had an IQ of 300, then the average would have to be more than $\frac{1}{3} \times 300 = 100$. So, the probability that a randomly chosen person has an IQ of 300 or more is at most 1/3. By the same logic, we can also conclude that at most 2/3 of the population can have an IQ of 150 or more.

Of course, these are not very strong conclusions. No IQ of over 300 has ever been recorded; and while many IQ's of over 150 have been recorded, the fraction of the population that actually has an IQ that high is very much smaller than 2/3. But though these conclusions are weak, we reached them just using the fact that the average IQ is 100 — along with another fact we took for granted: IQ is never negative. Using only these facts, we can't derive smaller fractions, because there are nonnegative random variables with mean 100 that achieve these fractions. For example, if we choose a random variable equal to 300 with probability 1/3 and 0 with probability 2/3, then its mean is 100, and the probability of a value of 300 or more really is 1/3.

Theorem 19.1.1 (Markov's Theorem). *If R is a nonnegative random variable, then for all $x > 0$*

$$\Pr[R \geq x] \leq \frac{\mathbb{E}[R]}{x}$$

Our focus is deviation from the mean, so it's useful to rephrase Markov's Theorem this way:

Corollary 19.1.2. *If R is a nonnegative random variable, then for all $c \geq 1$*

$$\Pr[R \geq c \times \mathbb{E}[R]] \leq \frac{1}{c}$$

This corollary follows immediately from Markov's Theorem by letting $x = c \times \mathbb{E}[R]$.

1.1. Applying Markov's Theorem

(A reference to § 18.5.2 of the textbook is made here.)

In the hat-check problem, we ask what the probability is that x or more men get the right hat, i.e. what the value of $\Pr[G \geq x]$ is.

We can compute an upper bound with Markov's Theorem. Since we know $\mathbb{E}[G] = 1$, Markov's Theorem implies

$$\Pr[G \geq x] \leq \frac{\mathbb{E}[G]}{x} = \frac{1}{x}$$

For example, there is no greater than a 20% chance that 5 men get the right hat, regardless of the number of people at the dinner party.

The Chinese Appetizer problem is similar to the Hat-Check problem. In this case, n people are eating different appetizers arranged on a lazy susan. Someone spins the tray so that each person receives a random appetizer. What is the probability that everyone gets the same appetizer as before?

There are n equally likely orientations for the tray after it stops spinning. Everyone gets the right appetizer in just one of these n orientations. Therefore, the correct answer is $\frac{1}{n}$. But what probability do we get from Markov's Theorem? Let the random variable, R, be the number of people that get the right appetizer. Then of course, $\mathbb{E}[R] = 1$, so applying Markov's Theorem, we find:

$$\Pr[R \geq n] \leq \frac{\mathbb{E}[R]}{n} = \frac{1}{n}$$

In this rare case, Markov's Theorem is precisely correct. However, in most other cases, Markov's Theorem will not be so accurate. For example, it gives the same $\frac{1}{n}$ bound for the probability that everyone gets their own hat back in the Hat-Check problem, where the probability is atually $\frac{1}{n!}$