Welcome back to my random tidbits file! When I come up with interesting problems, I will put them here.

1 Probability Distributions and Weight Loss

I was keeping track of my own weight when I realized that my scale was sufficiently inconsistent that my weight loss was dominated by the statistical noise. So then I was curious what the best way of mitigating this is, mean or median of multiple measurements. One would suspect it's the mean, or one would know simply by having taken any real statistics class, but I'm curious.

1.1 Mean-based averaging

This one is easy. Assume we have n iid variables X_i with mean μ and variance σ^2, then the random variable corresponding to their average $\langle X_i \rangle$ has mean μ and variance $\frac{\sigma^2}{n}$, so standard deviation $\frac{\sigma}{\sqrt{n}}$. Thus, we have an unbiased estimator of the true mean and a variance that falls off like $\sim n^{-1/2}$.

1.2 Median-based averaging

This one is a bit more fun. Let's start with $n = 3$, then defining $f(x)$ the probability density function and $F_X(x) = f_X(X \le x)$ the cumulative distribution function, the probability density of the median $f_\eta(y)$ is given

$$f_\eta(y) = 6 f_X(y) F_X(y) \left(1 - F_X(y)\right) \tag{1}$$

the probability we choose one value greater than y the median and one less, multiplied by 6 for orderings. This seems to be a bit difficult to verify to be normalized in the general case, or that

$$\int\limits_{-\infty}^{\infty} f_\eta(y)\,\mathrm{d}y = \int\limits_{-\infty}^{\infty} \left[6 f_X(y) \int\limits_{-\infty}^{y} f_X(\xi)\,\mathrm{d}\xi \int\limits_{y}^{\infty} f_X(\zeta)\,\mathrm{d}\zeta \right] \mathrm{d}y = 1 \tag{2}$$

Let's just verify this in the uniform distribution case, and leave the general case as an exercise to brighter colleagues. We consider the normalized uniform distribution $f_X(x) = 1, x \in [0,1]$, or $F_X(x) = x, x \in [0,1]$. We confirm that the expression for f_η is normalized:

$$\int\limits_{0}^{1} 6y(1-y)\,\mathrm{d}y = 1 \tag{3}$$

We then wish to examine whether $f_\eta(y)$ is an unbiased estimator of μ. Again, we begin with examining a sub-case, where $f_X(x)$ is symmetric about its mean μ. This yields that $F_X(\mu) = 0.5$ and is odd about μ[1] and so that $F_X(y)\left(1 - F_X(y)\right)$ is also even/symmetric about μ. Finally, this implies that $f_\eta(y)$ as defined in Equation 1 is also symmetric about μ and we are done.

However, this analysis breaks down in the asymmetric case. We see that $F_X(y)\left(1 - F_X(y)\right)$ is *always* symmetric about the median η of f_X, since $F_X(\eta) = 0.5$. In general, the mean and median of a probability distribution are not equal, so there is no guarantee that $\langle f_\eta(y) \rangle = \langle f(y) \rangle$, and indeed we can verify for some contrived probability distribution such as

$$f_X(x) = \begin{cases} 2 & 0 \le x \le 0.25 \\ 1 & 0.5 \le x \le 1 \\ 0 & \text{else} \end{cases} \tag{4}$$

that $\langle f_X(x) \rangle = 0.4375$ while

$$\begin{aligned} \langle f_\eta(y) \rangle &= \int\limits_{0}^{0.25} 24y^2(1-2y)\,\mathrm{d}y + \int\limits_{0.5}^{1} 6y^2(1-y)\,\mathrm{d}y \qquad (5) \\ &\approx 0.4218 \qquad (6) \end{aligned}$$

[1] This is a slight abuse of terminology: we mean that $F_X(x-\mu) - 0.5 = -(F_X(-(x-\mu)) - 0.5)$.

1.3 Open Questions

- If we have discretized measurements, what are the statistics of mode-based averaging?
- Did I actually normalize the median-based averaging correctly, for a general probability distribution?