Statistics 243: *class notes*

William J. De Meo

9/26/97

1 Algorithms for Mean & Variance

Definition 1.1 The sample mean, $\bar{x}$, of a sample of n numbers is

$$\bar{x} = \frac{1}{n}\sum_i x_i$$

Definition 1.2 The sample variance, s^2, is

$$s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$$

These are not the implementations we would use on the computer.

$$\begin{aligned}
s^2 &= \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 \\
&= \frac{1}{n-1}\sum_i (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\
&= \frac{1}{n-1}\left[\sum_i x_i^2 - 2\bar{x}\sum_i x_i + n\bar{x}^2\right] \\
&= \frac{1}{n-1}\left[\sum_i x_i^2 - n\bar{x}^2\right]
\end{aligned}$$

However, if both $\sum x_i^2$ and $n\bar{x}^2$ are large and nearly equal, we may lose all the significant digits using this computation. We need to know when this will be a problem. Let Q be the numerator of s^2. Since $\sum_i x_i^2$ is roughly equal to $n\bar{x}^2$, we have $\sum_i x_i^2/n\bar{x}^2$ is roughly equal to 1.

$$\frac{Q + n\bar{x}^2}{n\bar{x}^2} \approx 1$$

$$\frac{Q}{n\bar{x}^2} + 1 \approx 1$$

If CV denotes std.dev/mean then $(CV)^2 + 1 \approx 1$
When CV^2 gets so small that adding to 1 doesn't change 1, we have a problem.

1.1 Method of Provisional Means (BMDP)

At each stage, recenter the data by subtracting the current estimate of the mean.

Definition 1.3

$$\bar{x}^{(k)} = \frac{1}{k}\sum_i x_i$$

$$S^{(k)} = \sum_i (x_i - \bar{x}^{(k)})^2$$

$$\bar{x}^{(k)} = \bar{x}^{(k-1)} + \frac{1}{k}(x_k - \bar{x}^{(k-1)})$$

$$s^{(k)} = s^{(k-1)} + (x^k - \bar{x}^{(k-1)} \dots$$

1.2 Method of Subtracting the 1st Observation

Calculate:

$$\sum (x - x_1)$$

(n extra flops) and

$$\sum (x - x_1)^2$$

(n extra flops)
Then use the desk calculator algorithm. To get the right mean, we need to add x_1 back, but the variance is still right.[1]

2 Random Number Generation

We often want to know the property of a statistic when assumptions are violated.

Next time we'll look at generation of uniform r.v.s, some other distributions, and some tests of randomness.

[1] Find a data set where this approach fails.