Some Notes on Statistics

Carl W. Akerlof
January 7, 2008

Probability Distribution Functions:

\[ \frac{dp}{dx} = f(x) \ ; \int_{x_{\text{min}}}^{x_{\text{max}}} f(x)dx = 1 \]

Cumulative distribution function:

\[ F(x) = \int_{x_{\text{min}}}^{x} f(x')dx' \]
\[ F(x_{\text{min}}) = 0 \ ; F(x_{\text{max}}) = 1 \]

Population statistics:

Mean: \[ \langle x \rangle = \int_{x_{\text{min}}}^{x_{\text{max}}} x f(x) \, dx \]
Variance: \[ \sigma^2_x = \langle x^2 \rangle - \langle x \rangle^2 = \int_{x_{\text{min}}}^{x_{\text{max}}} x^2 f(x)dx - \langle x \rangle^2 \]
Standard Deviation: \[ \sigma_x = \sqrt{\sigma^2_x} \]
Sum of variables: \[ x = \sum_{i=1}^{n} x_i \]
Mean: \[ \langle x \rangle = \sum_{i=1}^{n} \langle x_i \rangle \]
Variance: \[ \sigma^2_x = \sum_{i=1}^{n} \sigma^2_{x_i} \]

Sample Statistics:

Sample Mean: \[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]
Sample Variance: \[ \sigma^2_x = \frac{1}{n-1} \left( \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} x_i \right)^2 \right) \]
Sample Standard Deviation: \[ \sigma_x = \sqrt{\sigma^2_x} \]
Mean of Sample Means: \[ \langle \bar{x} \rangle = \langle x \rangle \]
Mean of Sample Variances: \[ \langle \sigma^2_x \rangle = \sigma^2_x \]
Variance of Sample Means: \[ \sigma^2_{\bar{x}} = \frac{1}{n} \sigma^2_x \]
Variance of Sample Variances:

\[
\sigma^2_{\sigma^2} = \frac{1}{n} \left( \frac{1}{n} \right) - \frac{4}{n} \left( \frac{1}{n} \right)^2 - \frac{n-3}{n(n-1)} \left( \frac{1}{n} \right)^2 \\
+ \frac{4(2n-3)}{n(n-1)} \left( \frac{1}{n} \right)^2 - \frac{2(2n-3)}{n(n-1)} \left( \frac{1}{n} \right)^4
\]

For Gaussian distributions: 
\[
f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

Variance of sample variances: \(\sigma^2_{\sigma^2} = \frac{2\sigma^4}{n-1}\)

Standard deviation of sample variance: \(\sigma_{\sigma^2} = \sqrt{\frac{2}{n-1}} \sigma^2\)

Variance of sample standard deviations: \(\sigma^2_{\sigma_s} = \frac{\sigma^2}{2(n-1)}\)

Standard deviation of sample standard deviations: \(\sigma_{\sigma_s} = \frac{\sigma}{\sqrt{2(n-1)}}\)

Variance of median: \(\sigma^2_{x_{1/2}} = \frac{1}{4nf^2(x_{1/2})}; F(x_{1/2}) \equiv \frac{1}{2}\)

For Gaussian distribution: \(\sigma^2_{x_{1/2}} = \frac{\pi}{2n} \sigma^2\)

Robust estimate of dispersion: Inter-Quartile Difference (IQD)

Assume \(F(x_{1/4}) \equiv \frac{1}{4}; F(x_{3/4}) \equiv \frac{3}{4}\)

Then: \(IQD = x_{3/4} - x_{1/4}\)

For a Gaussian distribution:

\[
\frac{IQD}{q^*} = \sigma \text{ where } q^* = 1.3489798; \quad 1 = \frac{1}{2} = \frac{1}{\sqrt{2\pi}} \int_{-q^*/2}^{q^*/2} e^{-x^2} dx
\]

\[
\sigma^2_{(IQD/q^*)} = \frac{2\pi}{n} \frac{q^*/2}{q^*/2 - \sigma^4}
\]

\[
= 2.7209 \frac{2\sigma^4}{n}
\]