Combining Estimates

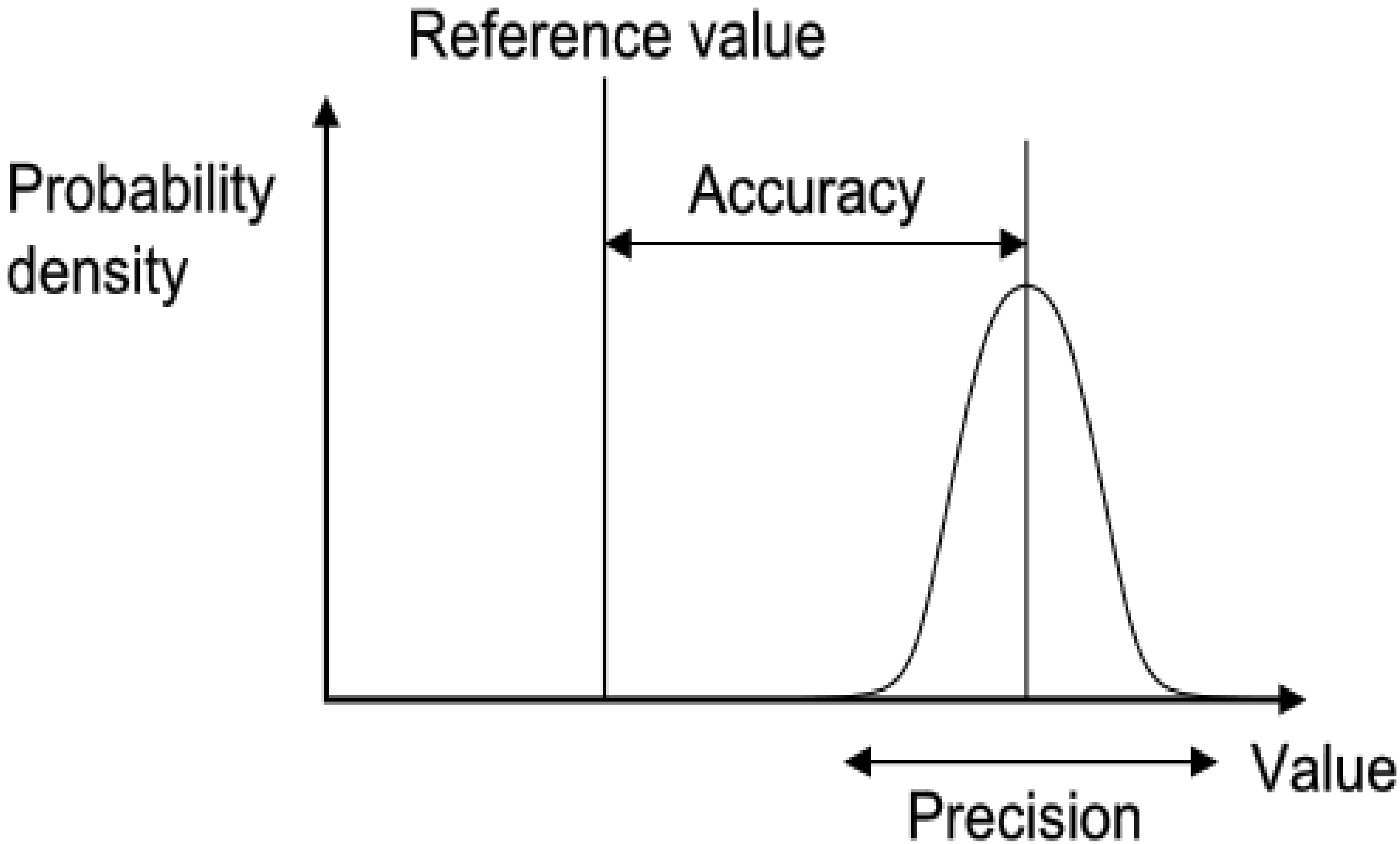
CLRS 2014
Tom Struppeck
The University of Texas at Austin

Goal:
Make a new estimator from old ones

Two cases:

- Case 1: Multiple estimates of one quantity
- Case 2: Desired quantity is a function of several quantities each with individual estimates

Accuracy vs Precision



Optimization

- Could optimize accuracy
- Could optimize precision
- Could try to optimize both simultaneously
  - Root mean square error

Two unbiased estimators for the same quantity

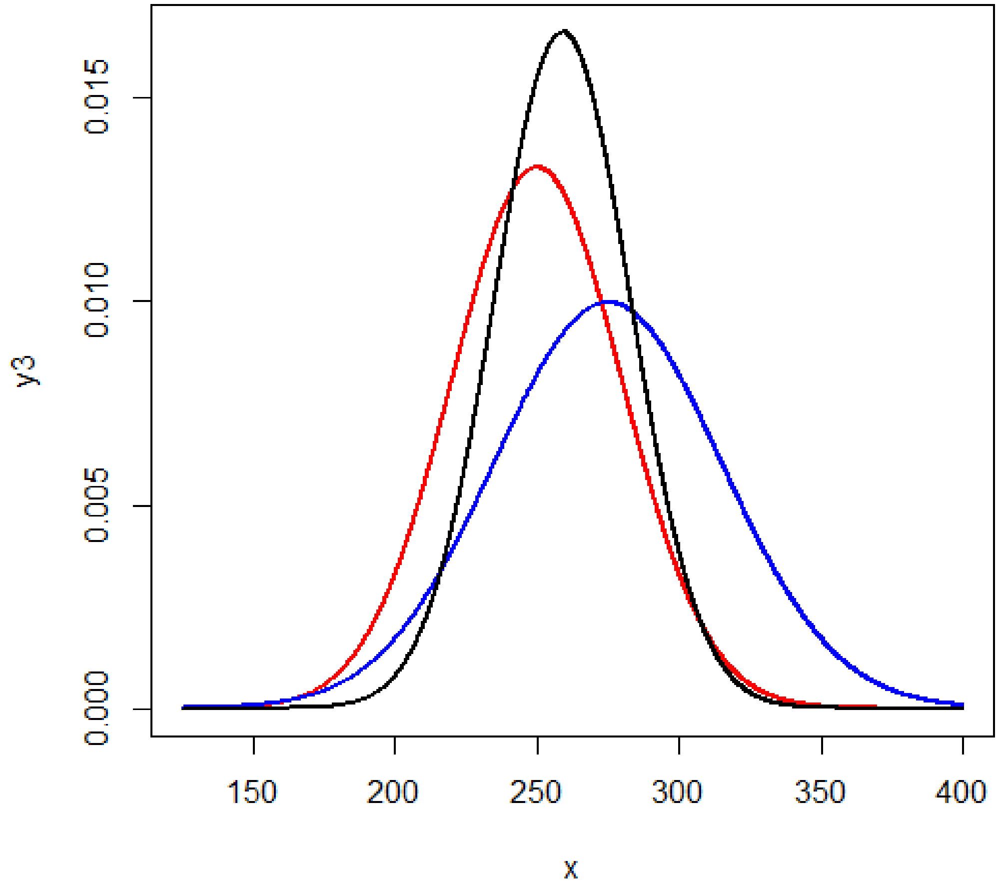
The **red** and **blue** curves are the densities for two estimators for an unknown parameter.

The **black** curve is the density of the optimal combination of these two estimators.

$Normal(250, 30^2)$

$Normal(275, 40^2)$

$Normal(259, 24^2)$



The optimal combination

Think of **red** as the original estimate and **blue** as new information.

black = Z* **blue** + (1 - Z)* **red**

Where:

$$Z = \frac{30^2}{30^2 + 40^2}$$

The new estimate is more precise

- Is it more accurate?

Here are the centers of the three distributions:

<................**250**.....**259**..........**275**............>

No matter where on the number line the true value is, 259 is closer to it than at least one of 250 and 275

Sums and Averages

Often, we are interested in a total. Typically we will have estimates for the summands.

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$$

This holds whether the summand are independent or not.

Divide by n to see it holds for averages, too.

The Central Limit Theorem

states that an average of independent random variables will converge (in distribution) to a normal random variable under very general conditions. (Assume IID for the rest of this slide.)

The variance of the n^{th} average will be proportional to $\frac{1}{n}$.

The variance of the n^{th} sum will be proportional to n.   NOTE: This goes to infinity with n.

What does that mean for us?

If we add together independent random variables, the variance of the sum is larger than the variance of the individual summands.

This is also true for correlated random variables unless the correlations are close to -1.

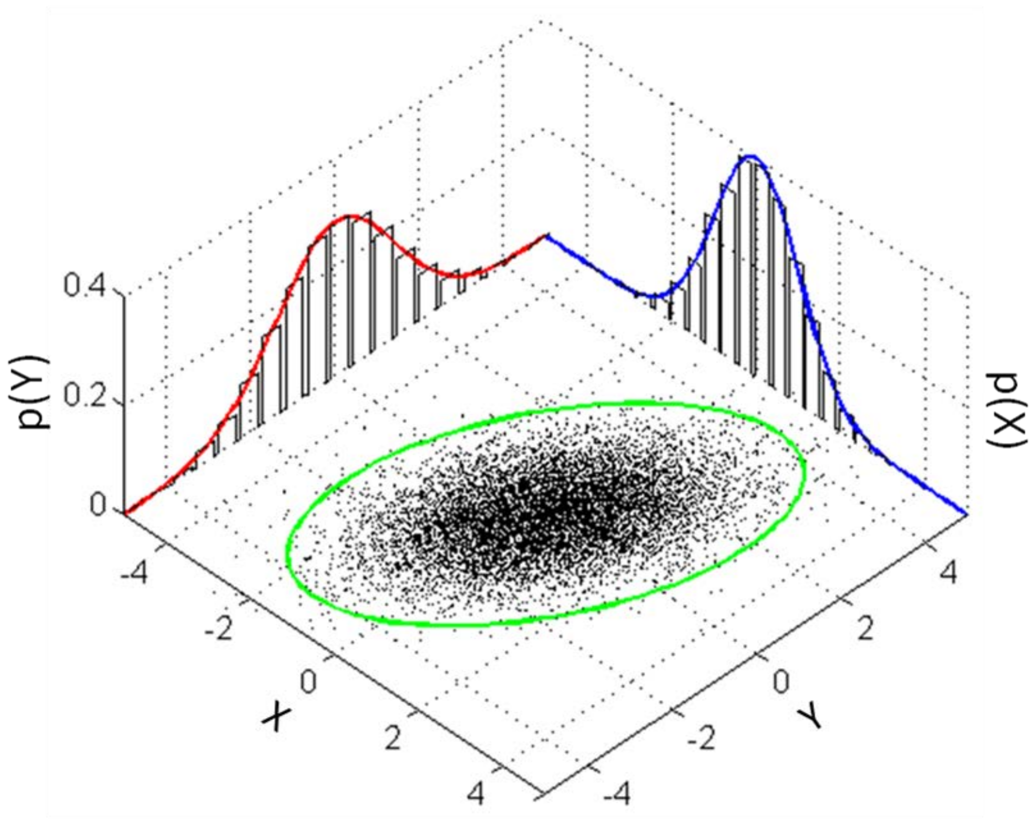
Our estimate of the sum will generally be less precise than our worst summand.

But all is not lost

Despite the fact that the variance is getting bigger, it is often the case that the expected value is also increasing, suppose that these are both increasing proportionally with n.

But, the standard deviation, being the square root of the variance, is growing more slowly, so the coefficient of variation (the ratio of the mean to the standard deviation) is going to 0.

An Example



An example with correlation

Line of Business	Expected losses	25th-percentile losses	75th-percentile losses	St. Dev. (Est.)	Estimated CV
A	100	90	110	14.8	0.148
B	225	150	300	111.2	0.494
C	350	200	500	222.4	0.635
Naïve Total	675	440	910	348.4	0.516
With Covariance Adjustment	675	465.3	884.7	310.9	0.461