ECE 307 – Techniques for Engineering Decisions

Using Data

George Gross

Department of Electrical and Computer Engineering

University of Illinois at Urbana-Champaign

FOCUS

- Use of historical data to obtain probability distributions
- The interpretation of probability information
- Use of estimators
- Application example

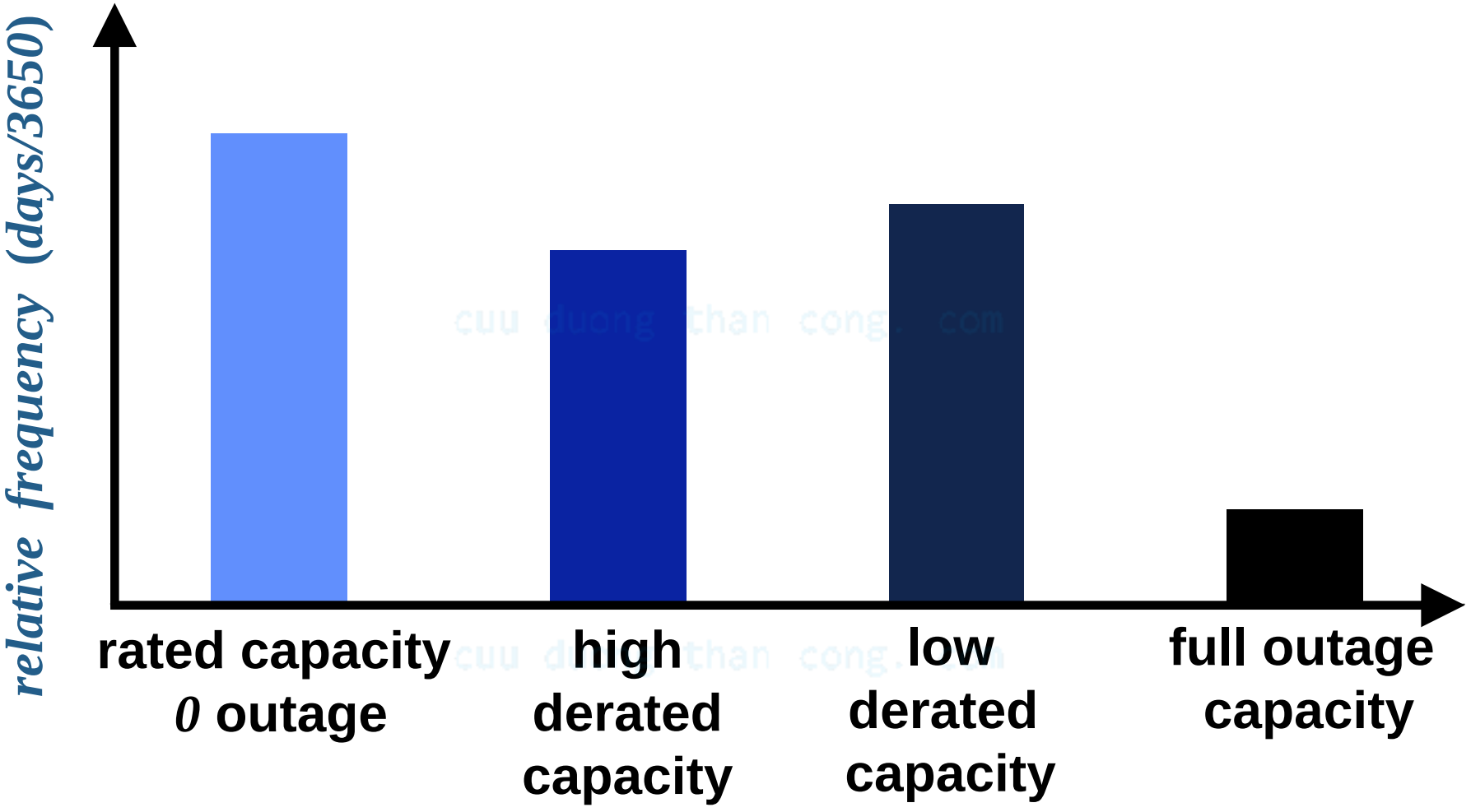
EXAMPLE

- Consider the interpretation of the statement

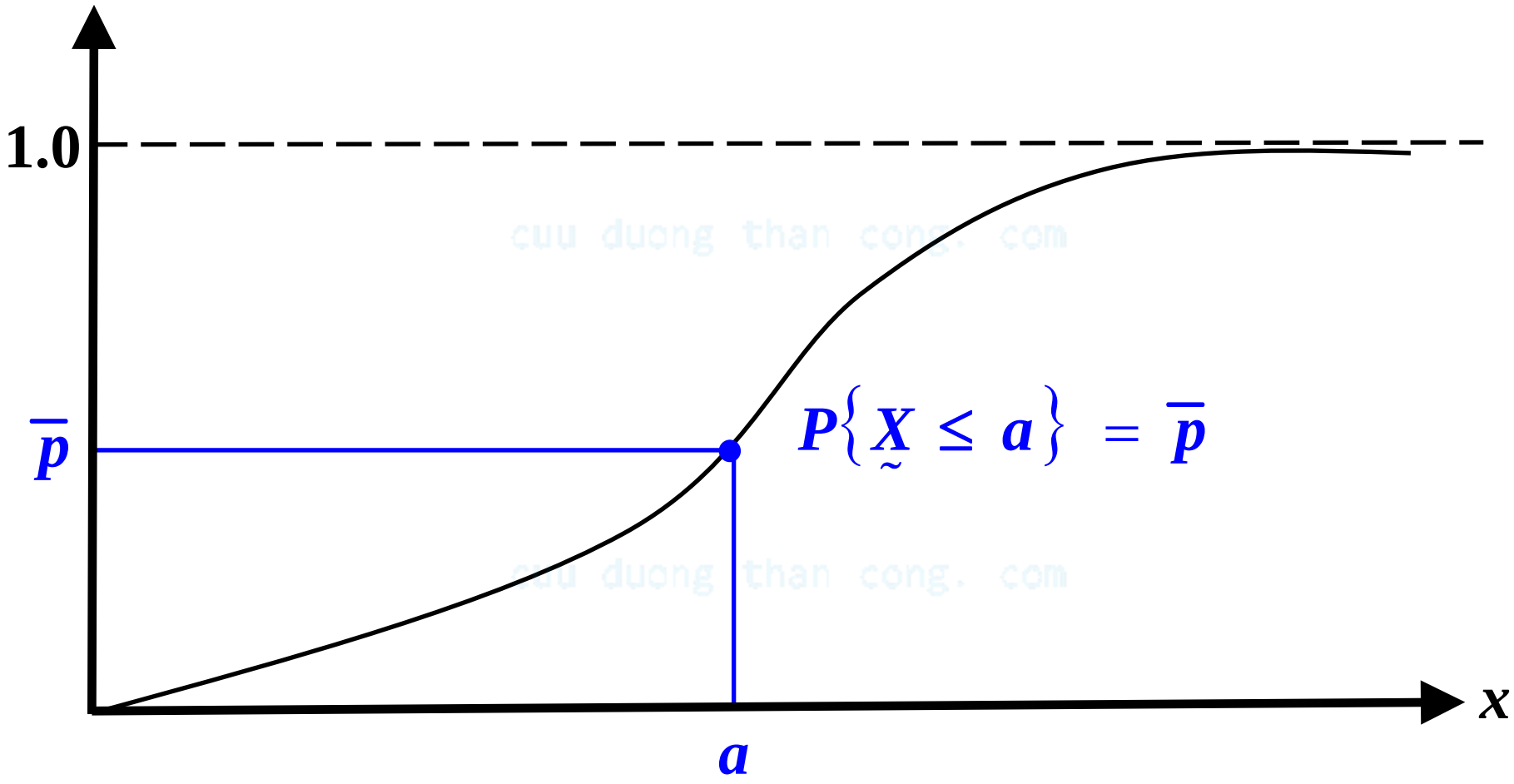
$$P\{sunny\ day\ in\ June\ in\ Champaign\} = 0.53$$

- June weather patterns in Champaign for the past 20 years are collected and every day is classified as either *sunny* or *not sunny*
- 600 days of June data are available with 318 or 53% of these days classified as sunny
- Given the long – term historical behavior, the probability of 0.53 makes sense

USE OF HISTOGRAMS



CONSTRUCTION OF THE *c.d.f.*



STATISTICAL PARAMETER ESTIMATORS

- **Estimator of the mean**

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

mean **of the distribution**

- **Estimator of the variance**

$$s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{n-1}$$

variance **of the distribution**

STATISTICAL PARAMETER ESTIMATORS

- We use a set of random samples $\{x_1, x_2, \ldots, x_n\}$ of a *r.v.* $\underset{\sim}{X}$: these are n randomly picked values from the sample space of $\underset{\sim}{X}$
- The estimator $\overline{x}$ computed with the set of random samples provides an estimate of

$$\mu = E\{\underset{\sim}{X}\}$$

- The estimator s^2 computed with the set of random samples provides an estimate of

$$\sigma^2 = var\{\underset{\sim}{X}\}$$

EXAMPLE: TACO SHELLS

- This application example focuses on taco shells and is concerned with the high breakage rate in the shipment of most taco shells: typical rate is $10 - 15\ \%$
- A company with a new shipping container claims to have a lower, approximately 5% breakage rate
- This company's price is $\$\ 25$ for a 500 – taco shell box vs. $\$\ 23.75$ for a 500 – taco shell box of the current supplier

EXAMPLE: TACO SHELLS

- A test run using 12 boxes from the new company and 18 boxes from the current company is performed and used for comparison purposes: in other words, we pick randomly $\{x_1, x_2, \ldots, x_{12}\}$ from the sample space of the *r.v.* $\underset{\sim}{X}$ describing the new company shells and $\{y_1, y_2, \ldots, y_{18}\}$ from the sample space of the *r.v.* $\underset{\sim}{Y}$ describing the current company shells
- The data of the useable shells from the two suppliers are tabulated

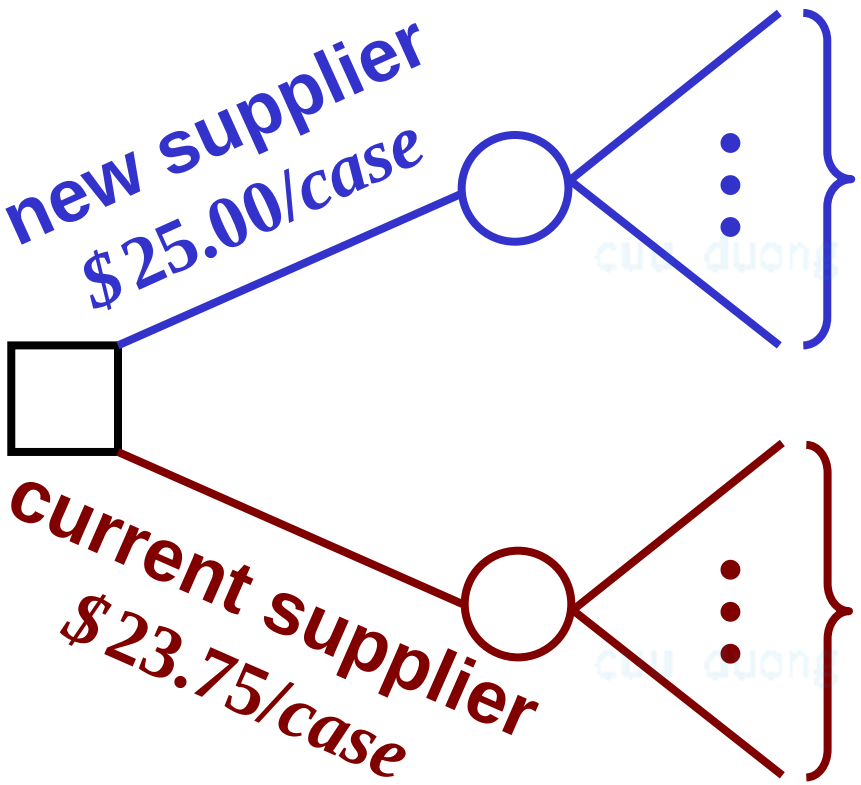
EXAMPLE: TACO SHELLS

useable shells

new supplier	
468	467
474	469
474	484
479	470
482	463
478	468

current supplier		
444	441	450
449	434	444
443	427	433
440	446	441
439	452	436
448	442	429

EXAMPLE: TACO SHELLS

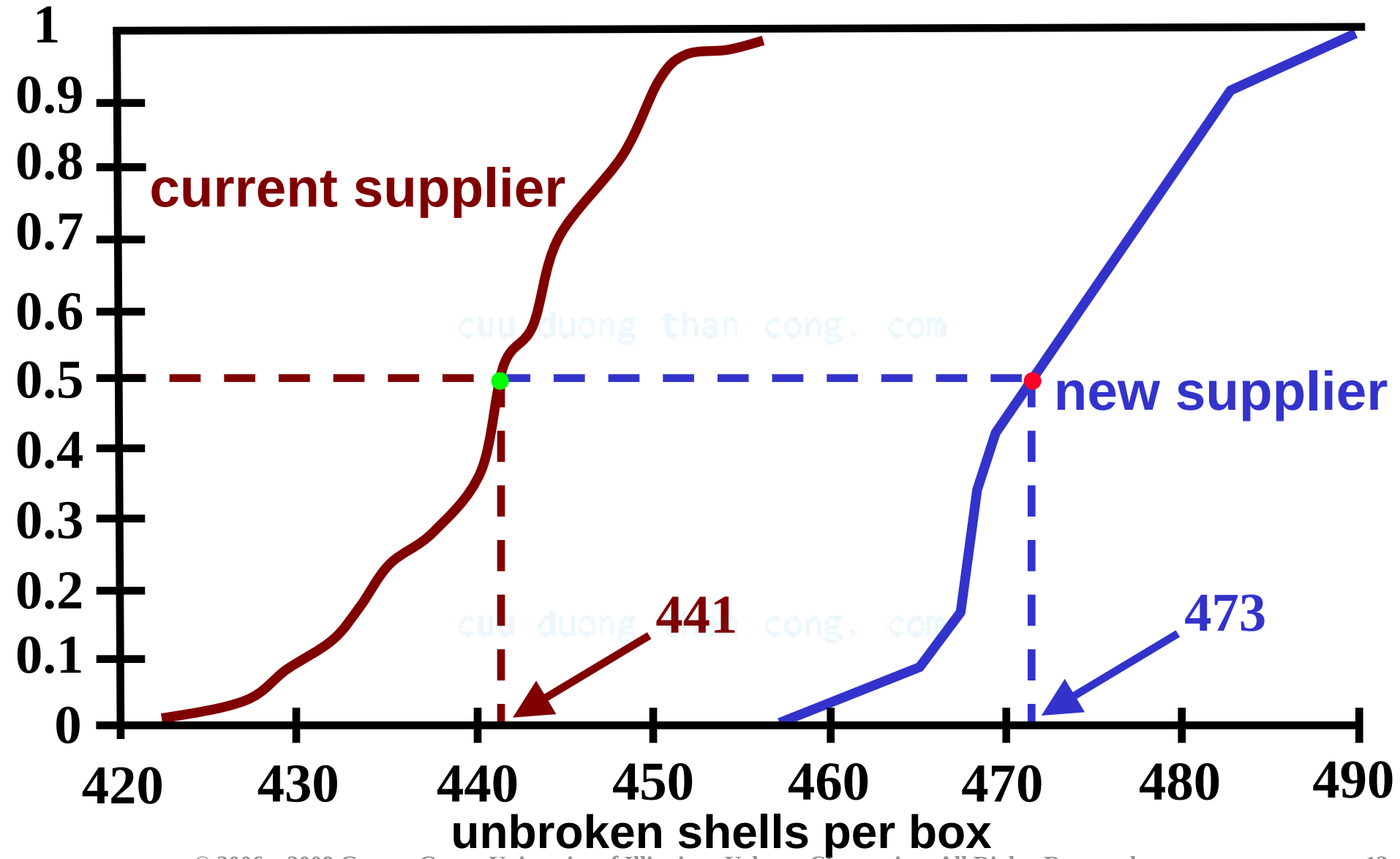


costs per unbroken shell

number of unbroken shells (x) $\frac{25}{x}$

number of unbroken shells (y) $\frac{23.75}{y}$

*c.d.f.*s CONSTRUCTED FOR THE TWO SUPPLIERS



*c.d.f.*s OF THE TWO SUPPLIERS

- Clearly, the new supplier has the higher expected number of useable shells per box; the two distributions, however, are highly similar

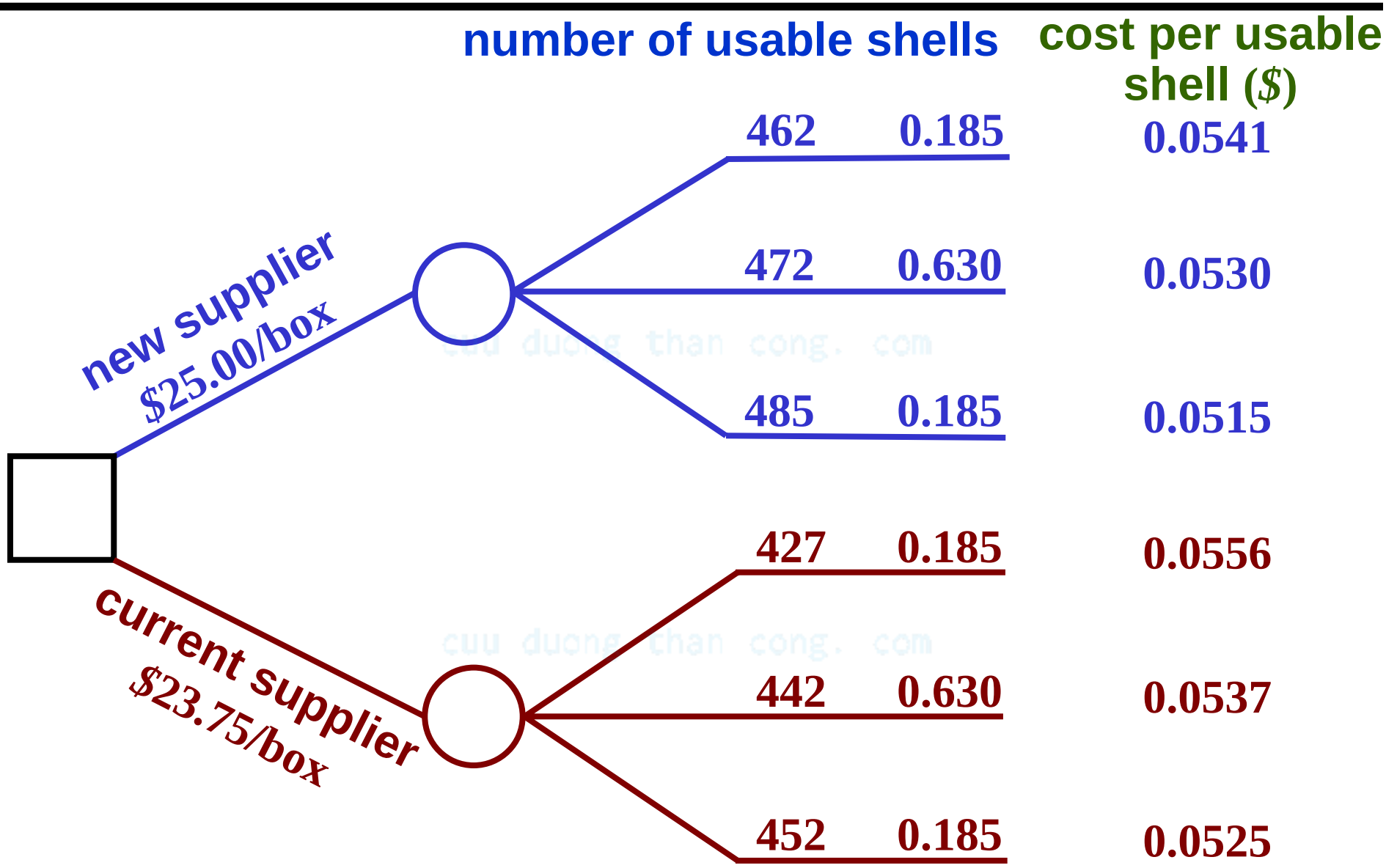
- The mean number of useable shells for the new supplier is 473 and so the expected costs per

*c.d.f.*s OF THE TWO SUPPLIERS

useable shell is $0.0529; the minimum (maximum) number of useable shells is 463(482)

- The mean number of useable shells for the current supplier is 441 and so the expected costs per useable shell is $0.0539; the minimum (maximum) number of useable shells is 429(452)

EXAMPLE: TACO SHELLS



COMMENTS

- We use the $c.d.f.$s to estimate the means of the two populations of suppliers

- Typically, the function

$$E\left\{\frac{1}{\underset{\sim}{X}}\right\} \neq \left[E\left\{\underset{\sim}{X}\right\}\right]^{-1}$$

COMMENTS

and so we cannot use the approximation

$$E\left\{\frac{25}{\underset{\sim}{X}}\right\} \approx \frac{25}{E\left\{\underset{\sim}{X}\right\}}$$

- This example demonstrates the usefulness of the *c.d.f.*s in applications even when they can only be approximated for the available data