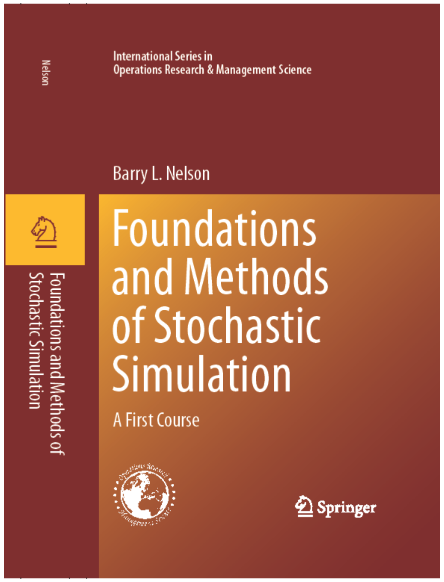


Chapter 7: Simulation Output

©Barry L. Nelson
Northwestern University
February 2013

Simulation output

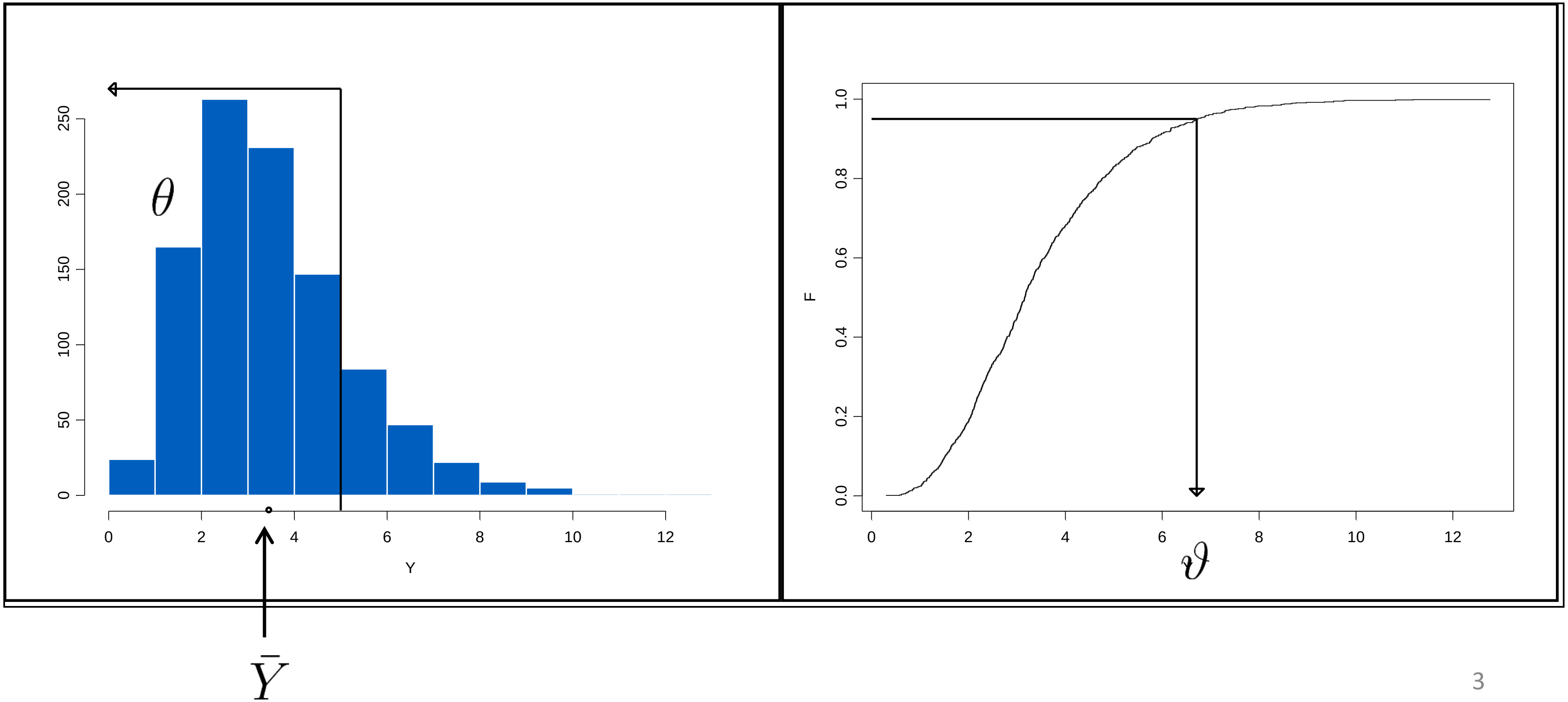
Suppose we make $n = 1000$ replications of Y, the time to complete the SAN. Let $Y_{(1)} \leq \cdots \leq Y_{(n)}$ be the *order statistics*.

What performance measures might be relevant?

- Mean time to complete the project, $\mu = \mathrm{E}(Y)$ estimated by the sample mean $\bar{Y} = \sum_{i=1}^{1000} Y_i / 1000$.
- Probability we complete the project in 5 days $\theta = F_Y(5)$, estimated by $\widehat{F}(5) = \#\{Y_i \leq 5\}/1000$.
- The 0.95 quantile $\vartheta = F_Y^{-1}(0.95)$, which is the date we can promise and be 95% sure of making it, estimated by $\widehat{F}^{-1}(0.95) = Y_{(950)}$.

Visualization

Visualizing means, probabilities and quantiles using the histogram and empirical cdf of 1000 SAN project completion times.



What measures are relevant?

For project planning, μ really does not make much sense.

The project will almost certainly not complete in exactly μ days, and it may not even be the most likely value.

If we think of the mean as the "long-run average," then it is most relevant *when the long-run average is what we will see* rather than a one-time outcome.

For the hospital information kiosk ($M/G/1$ queue), the long-run average waiting time is meaningful because the kiosk will serve many patients and visitors.

Q: Is $S(Y)$ relevant for the SAN? What about $S(Y)/\sqrt{1000}$?

Measures of error

No matter what performance measure we estimate, we need a *measure of error* (MOE) to establish how good it is.

Without an MOE, we cannot know if *any* of the digits in the estimate can be believed.

MOEs are also useful for experiment design: What number of replications and/or runlength is needed to attain an acceptable level of error?

Here we will talk about MOEs for i.i.d. (replication) data and do steady-state simulation (where run length matters) later.

95% CI for the SAN measures

A confidence interval is a measure of error; the wider it is the less certain we are about the true value.

Mean $\mathrm{E(Y)}$: $\mu \in 3.46 \pm 0.11$ days

Probability $F_Y(5) = \Pr\{Y \leq 5\}$: $\theta \in 0.17 \pm 0.02$

Quantile $F_Y^{-1}(0.95)$: $\vartheta \in [6.43, 7.05]$ days ($\widehat{\vartheta} = 6.71$)

The large sample CIs for μ and θ are justified by the CLT; the CI for ϑ is nonparametric.

Notice that we should not display more digits in the estimate than can be justified by the CI.

CI for the mean

$$\bar{Y} \pm z_{1-\alpha/2} \frac{S}{\sqrt{n}}$$

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \\ &= \frac{1}{n-1} \left[\sum_{i=1}^{n} Y_i^2 - \frac{1}{n} \left(\sum_{j=1}^{n} Y_j \right)^2 \right] \end{aligned}$$

The second expression allows S^2 to be computed in one pass through the data. Error decreases as $1/\sqrt{n}$.

CI for the probability

The estimator of θ is still a sample mean

$$\widehat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \le y)$$

Careful algebra gives

$$S^2 = \left(\frac{n}{n-1}\right)\widehat{F}(y)\left(1-\widehat{F}(y)\right)$$

When n is large the ratio $n/(n-1)$ is often treated as 1.

Remember that *relative error* is unbounded as $\theta \to 0$.

CI for the quantile

The q quantile $\vartheta = F_Y^{-1}(q)$ implies that $\Pr\{Y \leq \vartheta\} = q$.

Suppose we observe i.i.d. $Y_1, Y_2, \ldots, Y_n$. Then

$$\#\{Y_i \leq \vartheta\} \sim \text{Binomial}(n, q)$$

Therefore

$$\Pr\{Y_{(\ell)} \leq \vartheta\} = \Pr\{\text{at least } \ell \ Y_i\text{'s} \leq \vartheta\} = \sum_{i=\ell}^{n} \binom{n}{i} q^i (1-q)^{n-i}$$

To get a CI we look for $0 \leq \ell < u \leq n$ such that

$$\Pr\{Y_{(\ell)} \leq \vartheta < Y_{(u)}\} = \sum_{i=\ell}^{u-1} \binom{n}{i} q^i (1-q)^{n-i} \approx 1 - \alpha$$

Normal approximation

In general $\widehat{\vartheta} = \widehat{F}^{-1}(q) = Y_{(\lceil nq \rceil)}$ with CI $[Y_{(\ell)}, Y_{(u)}]$.

A large n normal approximation to the binomial gives approximations for ℓ and u:

$$\begin{aligned} \widehat{\ell} &= \left\lfloor nq - z_{1-\alpha/2}\sqrt{nq(1-q)} \right\rfloor \\ \widehat{u} &= \left\lceil nq + z_{1-\alpha/2}\sqrt{nq(1-q)} \right\rceil \end{aligned}$$

For $q = 0.95$, $n = 1000$ and $z_{0.975} = 1.96$, we have 95% confidence interval for the 0.95 quantile of $[Y_{(936)}, Y_{(964)}]$.

Notice that ℓ and u are completely independent of the data.

How hard is quantile estimation?

Just as probability estimation is relatively more difficult as $\theta \to 0$, we should expect extreme quantiles to be more difficult to estimate.

Although they are not means, sample quantiles do satisfy a CLT:

If F_Y is strictly increasing and has a density f_Y, then

$$\sqrt{n}\left(\widehat{\vartheta} - \vartheta\right) \xrightarrow{D} \mathrm{N}\left(0, \frac{q(1-q)}{\left[f_Y(\vartheta)\right]^2}\right)$$

Recall that $\widehat{\vartheta} = Y_{(\lceil nq \rceil)}$.

Example: Extreme quantiles

Suppose $f_Y(y) = \mathrm{e}^{-y}, y \geq 0$. Then $\vartheta = -\ln(1-q)$ so

$$\mathrm{se}\left(\widehat{\vartheta}\right) \approx \sqrt{\frac{q(1-q)}{n\left[\exp\left(\ln(1-q)\right)\right]^2}} = \sqrt{\frac{q}{n(1-q)}}.$$

This increases dramatically as q approaches 1.

For example, the standard error of $\widehat{\vartheta}$ for estimating the 0.99 quantile is roughly ten times larger than the standard error for estimating the the median for the same n.

Q: Why don't we use the CLT to get a CI for ϑ?

Proof sketch

$$\begin{aligned}\Pr\left\{\sqrt{n}(\widehat{\vartheta}-\vartheta)\le y\right\} &= \Pr\left\{\widehat{\vartheta}\le \vartheta + y/\sqrt{n}\right\}\\ &= \Pr\left\{\widehat{F}(\vartheta + y/\sqrt{n}) > q\right\}\end{aligned}$$

because $\widehat{\vartheta} = Y_{(\lceil nq\rceil)} \le y$ if and only if $\widehat{F}(y) > q$.

$$\Pr\left\{\widehat{F}(\vartheta + y/\sqrt{n}) > q\right\}$$

$$= \Pr\left\{\sqrt{n}\left(\widehat{F}(\vartheta + y/\sqrt{n}) - q_n\right) > \sqrt{n}(q - q_n)\right\}$$

where $q_n = F_Y(\vartheta + y/\sqrt{n})$.

LHS: As $n \to \infty$

$$\sqrt{n}\left(\widehat{F}(\vartheta + y/\sqrt{n}) - q_n\right) \to \sqrt{n}\left(\widehat{F}(\vartheta) - q\right) \xrightarrow{D} \mathrm{N}(0, q(1-q))$$

RHS: As $n \to \infty$

$$\sqrt{n}(q - q_n) = \frac{-y\left[F_Y(\vartheta + y/\sqrt{n}) - F_Y(\vartheta)\right]}{y/\sqrt{n}} \longrightarrow -y f_Y(\vartheta)$$

Combining we get

$$\Pr\left\{\frac{\mathrm{N}(0, q(1-q))}{f_Y(\vartheta)} < y\right\} = \Pr\left\{\mathrm{N}\left(0, \frac{q(1-q)}{[f_Y(\vartheta)]^2}\right) < y\right\}$$

Risk vs. Error

One of the biggest areas of confusion in statistics is the difference between risk and error.

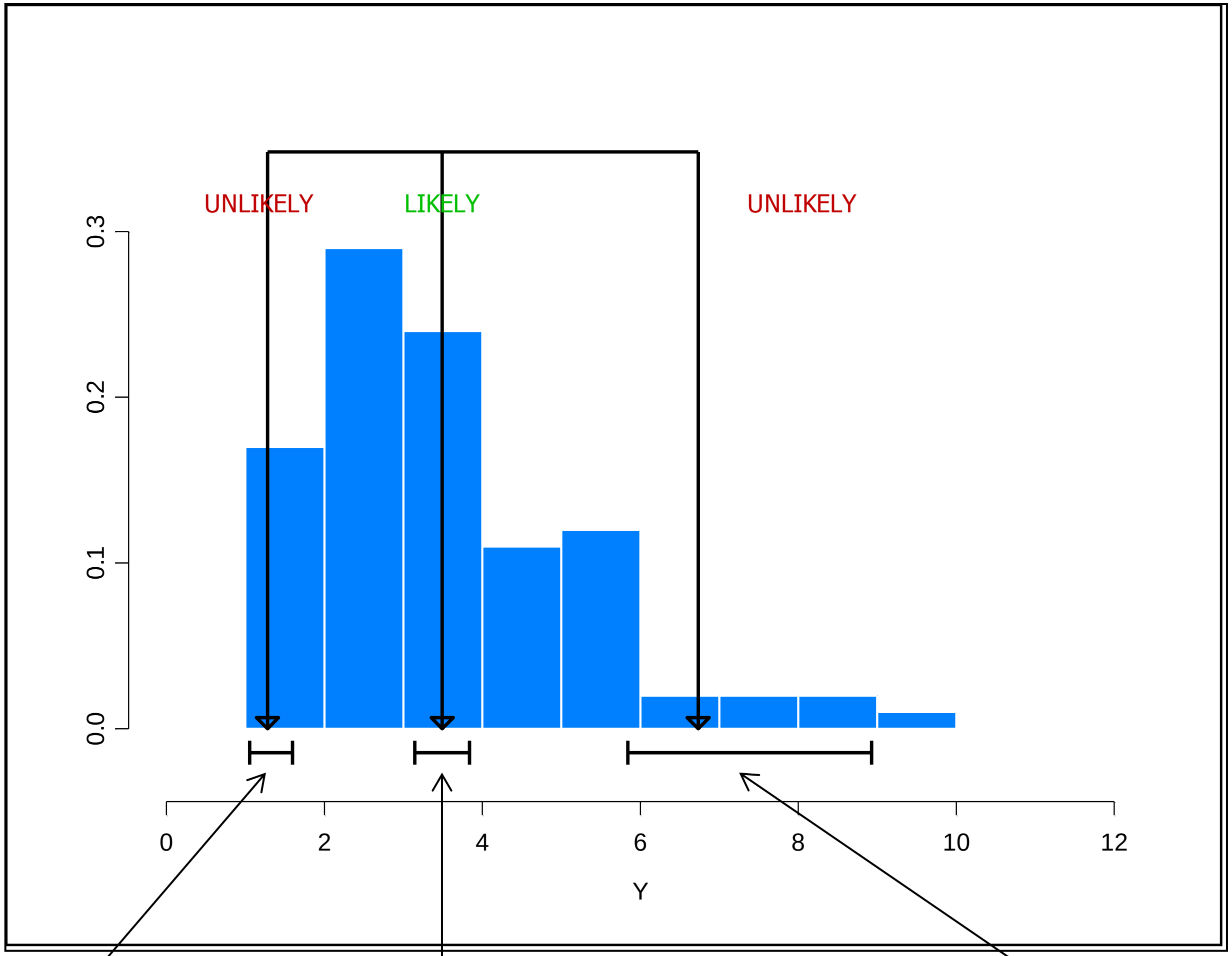
- *Measures of risk directly support decision making:* Should we bid this project, make this investment, deploy this system design?

  Risk is a property of the system that we cannot change by doing simulation.

- *Measures of error directly support experiment design.* Have we run enough simulation (e.g., replications) to be confident in our estimates of system performance?

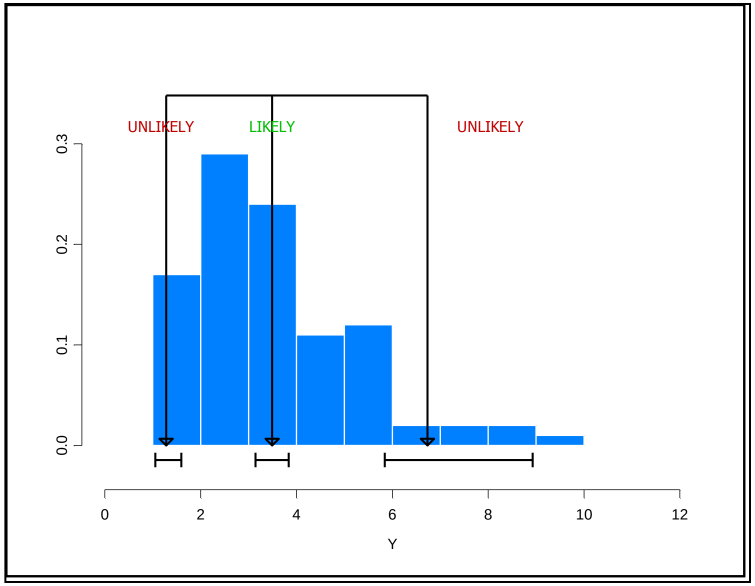
  Error is a property of the experiment which we can change by doing more or better simulation.

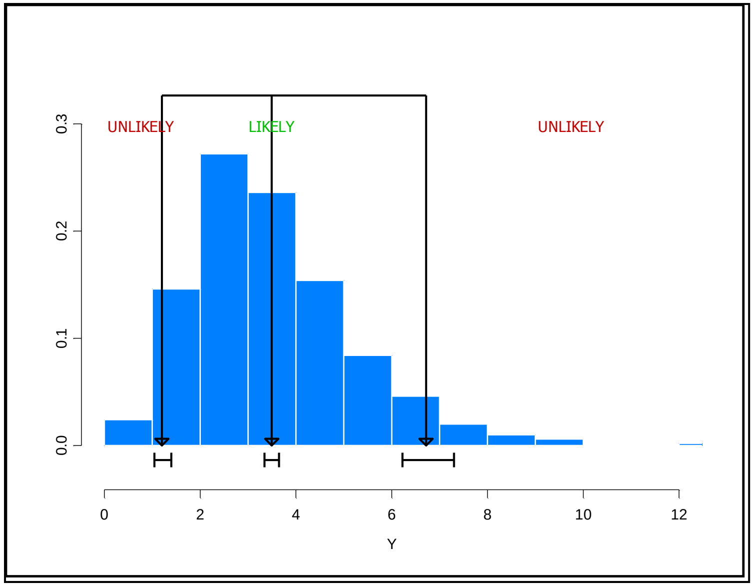
Measure of Risk and Error Plot



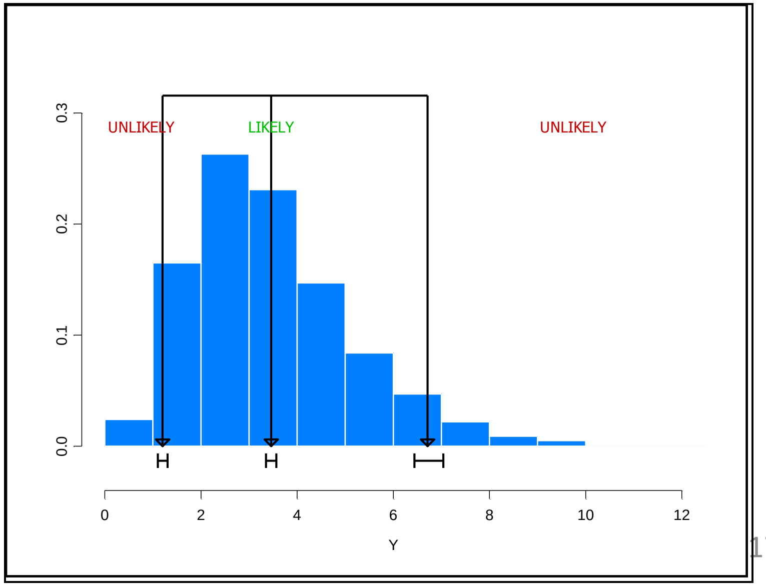
$Y_{(\lceil nq_1 \rceil)} \in [Y_{(\ell_1)}, Y_{(u_1)}]$ $\bar{Y} \pm zS/\sqrt{n}$ $Y_{(\lceil nq_2 \rceil)} \in [Y_{(\ell_2)}, Y_{(u_2)}]$

n = 100 → 500 → 1000 replications





We can simulate away error, but not risk since it is a property of the system. Changing risk requires changing the system.



Input uncertainty: What is it?

Consider an $M/M/\infty$ queue with arrival rate λ and mean service time τ, and let Y be the steady-state number of customers in the system.

Suppose λ and τ are not known, so we observe m i.i.d. interarrival times $A_1, A_2, \ldots, A_m$ and i.i.d. service times $X_1, X_2, \ldots, X_m$ from the "real world" and use them to fit input models:

$$\widehat{\lambda} = \left(\frac{1}{m} \sum_{i=1}^{m} A_i \right)^{-1}$$

$$\widehat{\tau} = \frac{1}{m} \sum_{i=1}^{m} X_i.$$

Stylized experiment

Simulate and record an observation of Y in steady state on each of n replications $Y_1, Y_2, \ldots, Y_n$.

Estimate the steady-state mean by the sample mean

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Then we can show that

$$\begin{aligned} \mathrm{E}(\bar{Y}) &= \frac{m}{m-1}\lambda\tau \\ \mathrm{Var}(\bar{Y}) &\approx \frac{\lambda\tau}{n} + \frac{2(\lambda\tau)^2}{m} \end{aligned}$$

Postmortem

- The need to estimate the input parameters introduces both bias and variance.
  - The bias diminishes quickly, and this is often the case.
  - But the variance due to "input uncertainty" can overwhelm the simulation variance.
- In a real problem we can't derive the effect.
- The impact is even more vexing if we don't know the model family, or we have no data.

Input uncertainty: What to do

Represent the output of the simulation on replication j, using estimated input distribution $\widehat{F}$, as

$$Y_j = \mu(\widehat{F}) + \varepsilon_j$$

where the $\{\varepsilon_j\}$ are i.i.d. $(0, \sigma_S^2)$ representing the simulation variability from replication to replication.

The mean term, $\mu(\widehat{F})$, depends on what input model we actually used in the simulation. Its variability, σ_I^2, represents input uncertainty.

Remark: We should expect σ_S^2 to depend on $\widehat{F}$ (why?), so this is clearly an approximation.

Idea: Bootstrap

1. Given real-world data $\{X_1, X_2, \ldots, X_m\}$, do:
2. For i from 1 to b
   (a) Generate the bootstrap sample $X_{i1}^\star, X_{i2}^\star, \ldots, X_{im}^\star$ by sampling m times with replacement from $\{X_1, X_2, \ldots, X_m\}$.
   (b) Fit $\widehat{F}_i^\star$ to $X_{i1}^\star, X_{i2}^\star, \ldots, X_{im}^\star$.
   If more than one input model, do Steps 2(a)–2(b) for each.
   (c) Simulate n replications $Y_{ij}, j = 1, 2, \ldots, n$ using input model(s) $\widehat{F}_i^\star$.
3. Estimate σ_I^2 using equations on the next slide.

Random-effects model

For a random effects model an estimator of σ_I^2 is

$$\widehat{\sigma}_I^2 = \frac{\widehat{\sigma}_T^2 - \widehat{\sigma}_S^2}{n}$$

where

$$\widehat{\sigma}_T^2 = \frac{n}{b-1} \sum_{i=1}^{b} \left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} \right)^2$$

and

$$\widehat{\sigma}_S^2 = \frac{1}{b(n-1)} \sum_{i=1}^{b} \sum_{j=1}^{n} \left(Y_{ij} - \bar{Y}_{i\cdot} \right)^2$$

Example M/M/∞

Suppose $\lambda = 5$, $\tau = 1$, we observe $m = 100$ real-world interarrival and service times, and make $n = 10$ replications. Then

$$\mathrm{Var}(\bar{Y}) \approx \frac{\lambda\mu}{n} + \frac{2(\lambda\tau)^2}{m} = \frac{5}{10} + \frac{50}{100} \approx \frac{\sigma_S^2}{n} + \sigma_I^2.$$

Running the procedure with $b = 100$ bootstrap samples gave $\widehat{\sigma}_S^2 = 5.321$ and $\widehat{\sigma}_I^2 = 0.546$.

Since $\widehat{\sigma}_S^2/10 = 0.5321$, we see that input uncertainty is approximately as large as estimation error.

M/M/∞: More details

We have $\{A_1, \ldots, A_{100}\}$ interarrival times and $\{X_1, \ldots, X_{100}\}$ service times from the "real world." Nominal simulation fits exponential distributions with $\widehat{\lambda} = 1/\bar{A}$ and $\widehat{\tau} = \bar{X}$ and makes n replications to get $\bar{Y}$.

Next we do the following $i = 1, 2, \ldots, b$ times:

Resample $\{A_1^\star, \ldots, A_{100}^\star\}$ and $\{X_1^\star, \ldots, X_{100}^\star\}$, fit exponential distributions with $\widehat{\lambda}^\star = 1/\bar{A}^\star$ and $\widehat{\tau}^\star = \bar{X}^\star$ and make n replications to get $Y_{ij}, j = 1, 2, \ldots, n$.

Analysis is based on $\{Y_{ij}\}$.

The good, the bad & the ugly

- This procedure is approximate, but easy to use; you don't even have to fit distributions since the simulation can be driven by the empirical distributions.
- What do we do with this information?
  - In many applications increasing *m* (amount of real-world data) is not possible. So all this tells you is how confident you can be in your results
  - If you *could* collect more real-world data, it does not indicate which input models account for the most uncertainty.