COMPARING SYSTEMS

Dave Goldsman

School of ISyE
Georgia Institute of Technology
Atlanta, GA, USA
sman@gatech.edu

July 29, 2018

Outline

Statistics / Simulation experiments are typically performed to analyze or compare a "small" number of systems, say ≤ 200.

The appropriate method depends on the type of comparison desired and properties of the output data.

If we analyze one system, we could use traditional confidence intervals (CIs) based on the normal or t-distributions from baby stats.

If we compare two systems, we could again use CIs from baby stats — maybe even clever ones based on paired observations.

If we compare > 2 systems, we may want to use *ranking and selection* techniques.

Confidence Intervals

Lots of possible confidence intervals:

- means, variances, quantiles.
- one-sample, two-sample cases (e.g., differences in means)

One-Sample Case:

- Interested in obtaining a two-sided $100(1-\alpha)\%$ CI for the unknown mean μ of a normal distribution.
- Suppose we have independent and identically distributed (i.i.d.) normal data $X_1, X_2, \ldots, X_n$.
- Assume unknown variance σ^2.
- Use the well-known t-distribution based CI, which I'll derive for your viewing pleasure.

First of all, recall that

- The sample mean $\bar{X}_n \equiv \sum_{i=1}^n X_i/n \sim \text{Nor}(\mu, \sigma^2/n)$.
- The sample variance $S_X^2 \equiv \sum_{i=1}^n (X_i - \bar{X}_n)^2/(n-1) \sim \sigma^2 \chi^2(n-1)/(n-1)$.
- $\bar{X}_n$ and S_X^2 are independent.

With these facts in mind, we have

$$T = \frac{\bar{X}_n - \mu}{\sqrt{S_X^2/n}} = \frac{\frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}}{\sqrt{S_X^2/\sigma^2}} \sim \frac{\text{Nor}(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} \sim t(n-1).$$

Letting the notation $t_{\gamma,\nu}$ denotes the $1-\gamma$ quantile of a t-distribution with ν degrees of freedom, we have

$$\begin{aligned} 1-\alpha &= P(-t_{\alpha/2,n-1} \leq T \leq t_{\alpha/2,n-1}) \\ &= P\left(-t_{\alpha/2,n-1} \leq \frac{\bar{X}_n - \mu}{\sqrt{S_X^2/n}} \leq t_{\alpha/2,n-1}\right) \\ &= P\left(\bar{X}_n - t_{\alpha/2,n-1} S_X/\sqrt{n} \leq \mu \leq \bar{X}_n + t_{\alpha/2,n-1} S_X/\sqrt{n}\right). \end{aligned}$$

So we have the following $100(1-\alpha)\%$ CI for μ,

$$\mu \in \bar{X}_n \pm t_{\alpha/2,n-1} S_X/\sqrt{n}.$$

Two-Sample Case: Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $\text{Nor}(\mu_X, \sigma_X^2)$ and $Y_1, Y_2, \ldots, Y_m$ are i.i.d. $\text{Nor}(\mu_Y, \sigma_Y^2)$.

A CI for the difference between $\mu_X - \mu_Y$ can be carried out by any of the following methods, all of which are from baby stats.

- pooled CI (use when σ_X^2 and σ_Y^2 are *equal but unknown*)
- approximate CI (use when σ_X^2 and σ_Y^2 are *unequal and unknown*)
- paired CI (use when $\text{Cov}(X_i, Y_i) > 0$)

In what follows, $\bar{X}$, $\bar{Y}$, S_X^2, and S_Y^2 are the obvious sample means and variances of the X's and Y's.

Pooled CI: If the X's and Y's are independent but with *common, unknown variance*, then the usual CI for the difference in means is

$$\mu_X - \mu_Y \in \bar{X} - \bar{Y} \pm t_{\alpha/2,n+m-2}\, S_P \sqrt{\frac{1}{n} + \frac{1}{m}},$$

where

$$S_P^2 \equiv \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

is the pooled variance estimator for σ^2.

Approximate CI: If the X's and Y's are independent but with arbitrary unknown variances, then the usual CI for the difference in means is

$$\mu_X - \mu_Y \;\in\; \bar{X} - \bar{Y} \pm t_{\alpha/2,\nu} \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}.$$

This CI is not quite exact, since it uses an *approximate* degrees of freedom,

$$\nu \;\equiv\; \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{(S_X^2/n)^2}{n+1} + \frac{(S_Y^2/m)^2}{m+1}} - 2.$$

Example: Times for people to parallel park two cars (assume normal).

A guy parks Honda X_i	Different (indep) guy parks Caddy Y_i
10	30
25	15
5	40
20	10
15	25

After a little algebra, we have

$$\bar{X} = 15, \quad \bar{Y} = 24, \quad S_X^2 = 62.5, \quad S_Y^2 = 142.5.$$

More algebra gives

$$\nu \;=\; \frac{6\Big(62.5+142.5\Big)^2}{(62.5)^2+(142.5)^2} - 2 \;=\; 8.4 \;\approx\; 8 \quad \text{(round down)}.$$

This yields the following 90% CI,

$$\mu_X - \mu_Y \;\in\; \bar{X} - \bar{Y} \pm t_{0.05,8}\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{n}} \;=\; -9 \pm 11.91,$$

which contains 0 and so is *inconclusive* about which of μ_X and μ_Y is bigger. $\square$

Paired CI: Again consider two competing normal pop'ns with unknown means μ_X and μ_Y. Suppose we collect observations from the two pop'ns in *pairs*.

Different pairs are *independent*, but the two obs'ns within the *same* pair may *not* be indep.

$$\text{indep} \begin{cases} \text{Pair } 1: & (X_1, Y_1) \\ \text{Pair } 2: & (X_2, Y_2) \\ \vdots & \vdots \\ \text{Pair } n: & \underbrace{(X_n, Y_n)}_{\text{not indep}} \end{cases}$$

Example: Think sets of twins. One twin takes a new drug, the other takes a placebo.

Idea: By setting up such experiments, we hope to be able to capture the difference between the two normal populations more precisely, since we're using the pairs to eliminate extraneous noise.

This will be the trick we use later on in this module when we use the simulation technique of *common random numbers*.

Here's the set-up. Take n pairs of observations:

$$\begin{aligned} X_1, X_2, \ldots, X_n &\overset{\text{iid}}{\sim} \text{Nor}(\mu_X, \sigma_X^2) \\ Y_1, Y_2, \ldots, Y_n &\overset{\text{iid}}{\sim} \text{Nor}(\mu_Y, \sigma_Y^2). \end{aligned}$$

(Technical assumption: All X_i's and Y_j's are jointly normal.)

We assume that the variances σ_X^2 and σ_Y^2 are *unknown* and possibly *unequal*.

Further, pair i is indep of pair j (between pairs), but X_i may not be indep of Y_i (within a pair).

Define the *pair-wise differences*, $D_i \equiv X_i - Y_i$, $i = 1, 2, \ldots, n$.

Then $D_1, D_2, \ldots, D_n \overset{\text{iid}}{\sim} \text{Nor}(\mu_D, \sigma_D^2)$, where $\mu_D \equiv \mu_X - \mu_Y$ (which is what we want the CI for), and

$$\sigma_D^2 \equiv \sigma_X^2 + \sigma_Y^2 - 2\text{Cov}(X_i, Y_i).$$

Idea: We hope that $\text{Cov}(X_i, Y_i)$ is pretty positive, which will result in lower σ_D^2 — low variance is a good thing!

Now the problem reduces to the old $\text{Nor}(\mu, \sigma^2)$ case with unknown μ and σ^2. So let's calculate the sample mean and variance as before.

$$\begin{aligned} \bar{D} &\equiv \frac{1}{n}\sum_{i=1}^{n} D_i \sim \text{Nor}(\mu_D, \sigma_D^2/n) \\ S_D^2 &\equiv \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2 \sim \frac{\sigma_D^2 \chi^2(n-1)}{n-1}. \end{aligned}$$

Just like before, get the CI

$$\mu_D \in \bar{D} \pm t_{\alpha/2,n-1}\sqrt{S_D^2/n}.$$

Example: Times for *the same person* to parallel park two cars.

Person	Park Honda	Park Cadillac	Difference
1	10	20	−10
2	25	40	−15
3	5	5	0
4	20	35	−15
5	15	20	−5

The individual people are indep, but the times for the same individual to park the two cars are not indep.

The 90% two-sided CI is therefore

$$\begin{aligned} \mu_D &\in \bar{D} \pm t_{0.05,4} \sqrt{S_D^2/n} \\ &= -9 \pm 2.13 \sqrt{42.5/5} = -9 \pm 6.21. \end{aligned}$$

This interval is entirely to the left of 0, indicating $\mu_D < 0$, i.e., Caddys take longer to park, on average. $\square$

This CI is quite a bit shorter (more informative) than the previous "approximate" two-sample CI, -9 ± 11.91, because the paired-t takes advantage of the correlation within observation pairs.

Moral: Use paired-t when you can.

Outline

One of the most important uses of simulation output analysis regards the comparison of competing systems or alternative system configurations.

Example: Evaluate two different "re-start" strategies that an airline can evoke following a major traffic disruption such as a snowstorm in the Northeast — which policy minimizes a certain cost function associated with the re-start?

Simulation is uniquely equipped to help the experimenter conduct this type of comparison analysis.

Many techniques: (i) classical CI's, (ii) variance reduction methods, and (iii) ranking and selection procedures.

Confidence Intervals for Mean Differences

With our airline example in mind, let $Z_{i,j}$ be the cost from the jth simulation replication of strategy i, $i = 1, 2$, $j = 1, 2, \ldots, b_i$.

Assume that $Z_{i,1}, Z_{i,2}, \ldots, Z_{i,b_i}$ are i.i.d. normal with unknown mean μ_i and unknown variance, $i = 1, 2$. Justification?...

(a) Get *independent* data by controlling the random numbers between replications.

(b) Get *identically distributed* costs between reps by performing the reps under identical conditions.

(c) Get approximately *normal* data by adding up (or averaging) many sub-costs to get overall costs for both strategies.

Confidence Intervals for Mean Differences

Goal: Obtain a $100(1-\alpha)\%$ CI for the difference in means, $\mu_1 - \mu_2$.

Suppose that the $Z_{1,j}$'s are independent of the $Z_{2,j}$'s and define

$$\bar{Z}_{i,b_i} \equiv \frac{1}{b_i}\sum_{j=1}^{b_i} Z_{i,j}, \quad i = 1, 2,$$

and

$$S_i^2 \equiv \frac{1}{b_i - 1}\sum_{j=1}^{b_i}(Z_{i,j} - \bar{Z}_{i,b_i})^2, \quad i = 1, 2.$$

Confidence Intervals for Mean Differences

An approximate $100(1-\alpha)\%$ CI is

$$\mu_1 - \mu_2 \in \bar{Z}_{1,b_1} - \bar{Z}_{2,b_2} \pm t_{\alpha/2,\nu}\sqrt{\frac{S_1^2}{b_1} + \frac{S_2^2}{b_2}}$$

where the (approx.) d.f. ν is given earlier in this module.

Suppose (as in airline example) that small cost is good.

- If the interval lies entirely to the left [right] of zero, then system 1 [2] is better.
- If the interval contains zero, then the two systems are, statistically, about the same.

Confidence Intervals for Mean Differences

Alternative strategy: Use a CI analogous to a paired-t test.

Take b replications from *both* strategies and set the difference $D_j \equiv Z_{1,j} - Z_{2,j}$ for $j = 1, 2, \ldots, b$.

Calculate the sample mean and variance of the differences:

$$\bar{D}_b \equiv \frac{1}{b}\sum_{j=1}^{b} D_j \text{ and } S_D^2 \equiv \frac{1}{b-1}\sum_{j=1}^{b}(D_j - \bar{D}_b)^2.$$

The $100(1-\alpha)\%$ paired-t CI is very efficient if $\text{Corr}(Z_{1,j}, Z_{2,j}) > 0$.

$$\mu_1 - \mu_2 \in \bar{D}_b \pm t_{\alpha/2,b-1}\sqrt{S_D^2/b}.$$

Common Random Numbers

Idea behind paired-t CI: Use *common random numbers*, i.e., use the same pseudo-random numbers in exactly the same ways for corresponding runs of each of the competing systems.

Example: Use the same customer arrival and service times when simulating different proposed configurations of a job shop.

By subjecting the alternative systems to identical experimental conditions, we hope to make it easy to distinguish which systems are best even though the respective estimators have sampling error.

Consider the case in which we compare two queueing systems, A and B, on the basis of their expected customer transit times, θ_A and θ_B — the smaller θ-value corresponds to the better system.

Suppose we have estimators $\hat{\theta}_A$ and $\hat{\theta}_B$ for θ_A and θ_B, respectively.

We'll declare A as the better system if $\hat{\theta}_A < \hat{\theta}_B$. If $\hat{\theta}_A$ and $\hat{\theta}_B$ are simulated independently, then the variance of their difference,

$$\mathrm{Var}(\hat{\theta}_A - \hat{\theta}_B) \;=\; \mathrm{Var}(\hat{\theta}_A) + \mathrm{Var}(\hat{\theta}_B),$$

could be very large; then our declaration might lack conviction.

If we could reduce $\mathrm{Var}(\hat{\theta}_A - \hat{\theta}_B)$, then we could be much more confident about our declaration.

CRN sometimes induces a high positive correlation between the point estimators $\hat{\theta}_A$ and $\hat{\theta}_B$.

Then we have

$$\begin{aligned} \mathrm{Var}(\hat{\theta}_A - \hat{\theta}_B) &= \mathrm{Var}(\hat{\theta}_A) + \mathrm{Var}(\hat{\theta}_B) - 2\mathrm{Cov}(\hat{\theta}_A, \hat{\theta}_B) \\ &< \mathrm{Var}(\hat{\theta}_A) + \mathrm{Var}(\hat{\theta}_B), \end{aligned}$$

and we obtain a savings in variance.

Demo Time! Queueing analysis. Exponential interarrival and service times. Which strategy yields shorter cycle times?

A. One line feeding into two parallel servers, or
B. Customers making a 50-50 choice between two lines each feeding into a single server?

Simulate each alternative for 20 replications of 1000 minutes.

The usual *independent* simulations of strategies A and B reveals gives a CI of $\mu_A - \mu_B \in -16.19 \pm 9.26$.

The use of *CRN* with the same arrival and service times across strategies gives $\mu_A - \mu_B \in -15.05 \pm 3.37$. Much tighter CIs! ☺

Antithetic Random Numbers

Opposite of CRN — Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are i.i.d. unbiased estimators for some parameter θ.

If we can induce *negative* correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$, then the average of the two is also unbiased and may have very low variance,

$$\begin{aligned}
\mathrm{Var}\left(\frac{\hat{\theta}_1 + \hat{\theta}_2}{2}\right) &= \frac{1}{4}\left[\mathrm{Var}(\hat{\theta}_1) + \mathrm{Var}(\hat{\theta}_2) + 2\mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2)\right] \\
&= \frac{1}{2}\left[\mathrm{Var}(\hat{\theta}_1) + \mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2)\right] \\
&< \frac{\mathrm{Var}(\hat{\theta}_1)}{2} \quad (\leftarrow \text{“usual” avg of two reps!}).
\end{aligned}$$

Example: Let's do some Monte Carlo integration, using ARN to obtain a nice variance reduction.

Consider the integral $I = \int_1^2 (1/x)\,dx$. (Because I have natural logger rhythm, I happen to know that the true answer is $\ell\mathrm{n}(2) \approx 0.693$.)

We'll use the following $n = 5$ $\mathrm{Unif}(0, 1)$ random numbers to come up with the usual estimator $\bar{I}_n$ for I:

$$0.85 \qquad 0.53 \qquad 0.98 \qquad 0.12 \qquad 0.45$$

Using the Monte Carlo integration notation from waaaay back in time with $g(x) = 1/x$,

$$\begin{aligned}
\hat{\theta}_1 = \bar{I}_n &= \frac{b-a}{n}\sum_{i=1}^{n} g(a+(b-a)U_i) \\
&= \frac{1}{5}\sum_{i=1}^{5} g(1+U_i) \\
&= \frac{1}{5}\sum_{i=1}^{5} \frac{1}{1+U_i} \\
&= 0.6563 \quad \text{(not bad).}
\end{aligned}$$

Now we'll use the following *antithetic* random numbers (which are all the "opposite" of the above PRNs, i.e., $1 - U_i$):

$$0.15 \qquad 0.47 \qquad 0.02 \qquad 0.88 \qquad 0.55$$

Then the antithetic version of the estimator is

$$\hat{\theta}_2 \;=\; \frac{1}{5}\sum_{i=1}^{5} \frac{1}{1+(1-U_i)} \;=\; 0.7475 \quad \text{(also not bad).}$$

But lookee here when you take the average of the two answers,

$$\frac{\hat{\theta}_1 + \hat{\theta}_2}{2} \;=\; 0.6989.$$

Wow — really close to the right answer! ☺

Control Variates

Another method to reduce estimator variance is related to regression.

Suppose that our goal is to estimate the mean μ of some steady-state simulation output process, $X_1, X_2, \ldots, X_n$. Suppose we somehow know the expected value of some other RV Y, and we also know that $\mathrm{Cov}(\bar{X}, Y) > 0$, where $\bar{X}$ is the sample mean.

Obviously, $\bar{X}$ is the "usual" estimator for μ. But let's look at another estimator for μ, namely, the *control-variate* estimator,

$$C = \bar{X} - \beta(Y - \mathrm{E}[Y]), \quad \text{for some constant } \beta.$$

Note that

$$\mathrm{E}[C] \;=\; \mathrm{E}[\bar{X}] - \beta(\mathrm{E}[Y] - \mathrm{E}[Y]) \;=\; \mathrm{E}[\bar{X}] \;=\; \mu.$$

Further,

$$\mathrm{Var}(C) \;=\; \mathrm{Var}(\bar{X}) + \beta^2\,\mathrm{Var}(Y) - 2\beta\,\mathrm{Cov}(\bar{X}, Y).$$

And then we can minimize $\mathrm{Var}(C)$ with respect to β. Differentiating,

$$\beta \;=\; \frac{\mathrm{Cov}(\bar{X}, Y)}{\mathrm{Var}(Y)}.$$

Thus,

$$\mathrm{Var}(C) \;=\; \mathrm{Var}(\bar{X}) - \frac{\mathrm{Cov}^2(\bar{X}, Y)}{\mathrm{Var}(Y)} \;<\; \mathrm{Var}(\bar{X}). \quad \square$$

Examples: We might try to estimate a population's mean weight μ using observed weights $X_1, X_2, \ldots$ with corresponding heights $Y_1, Y_2, \ldots$ as controls (assuming that $\mathrm{E}[Y]$ is known).

We could estimate the price of an American stock option (which is tough) using the corresponding European option price (which is easy) as a control.

In any case, many simulation texts give advice on how to run the simulations of the competing systems so as to use CRN, ARN, and control variates.

Outline

Ranking, selection, and multiple comparisons methods form another class of statistical techniques used to compare alternative systems.

Here, the experimenter is interested in selecting the best of a number (≥ 2) of competing processes.

Specify the desired probability of correctly selecting the best process, especially if the best process is significantly better than its competitors.

These methods are simple to use, fairly general, and intuitively appealing (see Bechhofer, Santner, and Goldsman 1995).

For > 2 systems, we could use methods such as simultaneous CIs and ANOVA. But those methods don't tell us much except that "at least one of the systems is different than the others", which is no surprise.

And what measures do you use to compare different systems?

- Which has the biggest mean?
- The smallest variance?
- The highest probability of yielding a success?
- The lowest risk?
- A combination of criteria?

Remainder of this module: We present ranking & selection procedures to find the best system with respect to one parameter.

Examples:

- Great Expectations: Which of 10 fertilizers produces the largest mean crop yield? (Normal)
- Great Expectorants: Find the pain reliever that has the highest probability of giving relief for a cough. (Binomial)
- Great Ex-Patriots: Who is the most-popular former New England football player? (Multinomial)

R&S selects the best system, or a subset of systems that includes the best.

- Guarantee a probability of a correct selection.
- Multiple Comparisons Procedures (MCPs) add in certain confidence intervals.

R&S is relevant in simulation:

- Normally distributed data by batching.
- Independence by controlling random numbers.
- Multiple-stage sampling by retaining the seeds.

Find the Normal Distribution with the Largest Mean

We give procedures for selecting that one of k *normal* distributions having the largest mean.

We use the *indifference-zone* approach.

Assumptions: Independent $Y_{i1}, Y_{i2}, \ldots$ $(1 \le i \le k)$ are taken from $k \ge 2$ normal populations $\Pi_1, \ldots, \Pi_k$. Here Π_i has *unknown* mean μ_i and *known* or *unknown* variance σ_i^2.

Denote the vector of means by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ and the vector of variances by $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_k^2)$.

The ordered (but unknown) μ_i's are $\mu_{[1]} \le \cdots \le \mu_{[k]}$.

The system having the largest mean $\mu_{[k]}$ is the "best."

Find the Normal Distribution with the Largest Mean

Goal: To select the population associated with mean $\mu_{[k]}$.

A *correct selection* (CS) is made if the Goal is achieved.

Indifference-Zone Probability Requirement: For specified constants $(P^\star, \delta^\star)$ with $\delta^\star > 0$ and $1/k < P^\star < 1$, we require

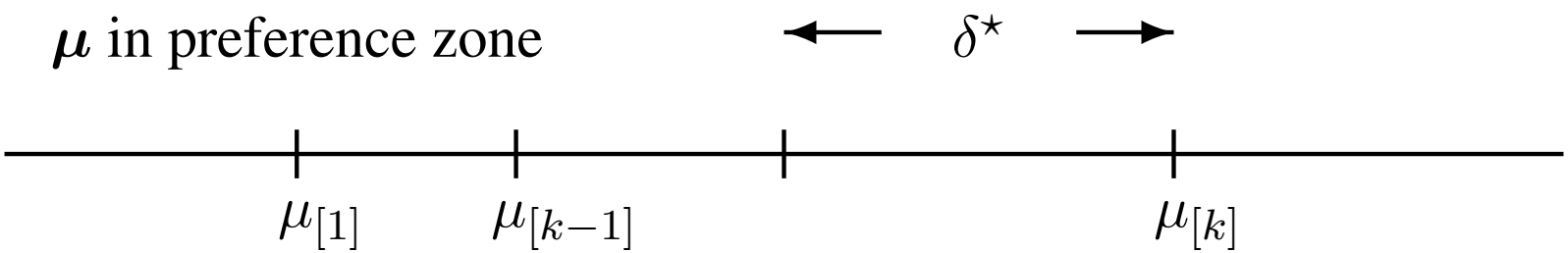
$$P(\text{CS}) \geq P^\star \quad \text{whenever} \quad \mu_{[k]} - \mu_{[k-1]} \geq \delta^\star. \tag{1}$$

The constant $\delta^\star$ can be thought of as the "smallest difference worth detecting."

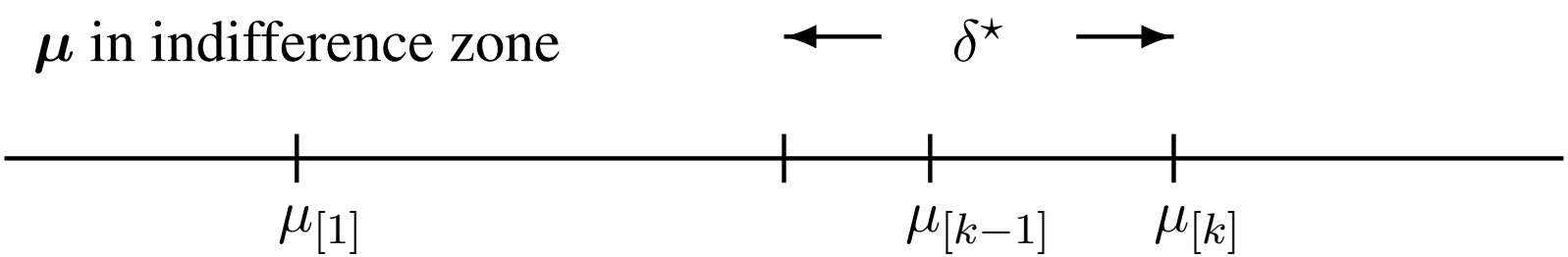
The probability in (1) depends on the differences $\mu_i - \mu_j$, the sample size n, and $\boldsymbol{\sigma}^2$.

Find the Normal Distribution with the Largest Mean

Parameter configurations $\boldsymbol{\mu}$ satisfying $\mu_{[k]} - \mu_{[k-1]} \geq \delta^\star$ are in the *preference-zone* for a correct selection.



If $\mu_{[k]} - \mu_{[k-1]} < \delta^\star$, then you're in the *indifference-zone*.



Find the Normal Distribution with the Largest Mean

Any procedure that guarantees (1) is said to be employing the indifference-zone approach.

There are 100's of such procedures. Highlights:

- Single-Stage Procedure (Bechhofer 1954)
- Two-Stage Procedure (Rinott 1979)
- Sequential Procedure (Kim and Nelson 2001)

Single-Stage Procedure $\mathcal{N}_{\mathrm{B}}$ (Bechhofer 1954)

This procedure takes all necessary observations and makes the selection decision at once (in a single stage).

Assumes populations have *common known variance*.

For the given k and specified $(P^\star, \delta^\star/\sigma)$, determine sample size n.

Take a random sample of n observations Y_{ij} $(1 \leq j \leq n)$ in a single stage from Π_i $(1 \leq i \leq k)$.

Find the Normal Distribution with the Largest Mean

Calculate the k sample means, $\bar{Y}_i = \sum_{j=1}^{n} Y_{ij}/n$ $(1 \leq i \leq k)$.

Select the population that yielded the largest sample mean, $\bar{Y}_{[k]} = \max\{\bar{Y}_1, \ldots, \bar{Y}_k\}$, as the one associated with $\mu_{[k]}$.

Very intuitive — all you have to do is figure out n.

- from a table (easy), or
- from a multivariate normal quantile (not too bad), or
- via a separate simulation (if all else fails)

Find the Normal Distribution with the Largest Mean

k	$P^\star$	$\delta^\star/\sigma$									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
2	0.75	91	23	11	6	4	3	2	2	2	1
	0.90	329	83	37	21	14	10	7	6	5	4
	0.95	542	136	61	34	22	16	12	9	7	6
	0.99	1083	271	121	68	44	31	23	17	14	11
3	0.75	206	52	23	13	9	6	5	4	3	3
	0.90	498	125	56	32	20	14	11	8	7	5
	0.95	735	184	82	46	30	21	15	12	10	8
	0.99	1309	328	146	82	53	37	27	21	17	14
4	0.75	283	71	32	18	12	8	6	5	4	3
	0.90	602	151	67	38	25	17	13	10	8	7
	0.95	851	213	95	54	35	24	18	14	11	9
	0.99	1442	361	161	91	58	41	30	23	18	15

Common Sample Size n per Population Required by $\mathcal{N}_{\mathrm{B}}$

Remark: Don't really need the above table. You can directly calculate

$$n \;=\; \left\lceil 2\left(\sigma Z_{k-1,1/2}^{(1-P^\star)}/\delta^\star\right)^2 \right\rceil,$$

where $\lceil\cdot\rceil$ rounds up, and the constant $Z_{k-1,1/2}^{(1-P^\star)}$ is an upper equicoordinate point of a certain multivariate normal distribution.

The value of n satisfies the Probability Requirement (1) for any $\boldsymbol{\mu}$ with

$$\mu_{[1]} \;=\; \mu_{[k-1]} \;=\; \mu_{[k]} - \delta^\star. \tag{2}$$

Configuration (2) is the *slippage* configuration (since $\mu_{[k]}$ is larger than the other μ_i's by a fixed amount). It turns out that for Procedure $\mathcal{N}_{\mathrm{B}}$, (2) is also the *least-favorable* (LF) configuration because, for fixed n, it minimizes the $P(\mathrm{CS})$ among all $\boldsymbol{\mu}$ in the preference-zone.

The next thing we'll do is to calculate n (without using multivariate normal tables).

The value of n will be the smallest value that satisfies the Probability Requirement when $\boldsymbol{\mu}$ is in the "worst-case" LF configuration.

We'll assume without loss of generality that Π_k has the largest μ_i.

And recall our old notation friends $\phi(\cdot)$ and $\Phi(\cdot)$ — the standard normal p.d.f. and c.d.f.

Find the Normal Distribution with the Largest Mean

$$
\begin{aligned}
P^\star &= P(\mathrm{CS}\,|\,\mathrm{LF}) = P\{\bar{Y}_i < \bar{Y}_k,\, i = 1, \ldots, k-1 \,|\, \mathrm{LF}\} \\
&= P\left\{ \frac{\bar{Y}_i - \mu_k}{\sqrt{\sigma^2/n}} < \frac{\bar{Y}_k - \mu_k}{\sqrt{\sigma^2/n}},\, i = 1, \ldots, k-1 \,\middle|\, \mathrm{LF} \right\} \\
&= \int_{\mathbb{R}} P\left\{ \frac{\bar{Y}_i - \mu_k}{\sqrt{\sigma^2/n}} < x,\, i = 1, \ldots, k-1 \,\middle|\, \mathrm{LF} \right\} \phi(x)\, dx \\
&= \int_{\mathbb{R}} P\left\{ \frac{\bar{Y}_i - \mu_i}{\sqrt{\sigma^2/n}} < x + \frac{\sqrt{n}\delta^\star}{\sigma},\, i = 1, \ldots, k-1 \right\} \phi(x)\, dx \\
&= \int_{\mathbb{R}} \Phi^{k-1}\left(x + \frac{\sqrt{n}\delta^\star}{\sigma} \right) \phi(x)\, dx \;=\; \int_{\mathbb{R}} \Phi^{k-1}(x+h)\phi(x)\, dx.
\end{aligned}
$$

Now solve numerically for h, and then set $n = \lceil (h\sigma/\delta^\star)^2 \rceil$.

Find the Normal Distribution with the Largest Mean

Example: Suppose $k = 4$ and we want to detect a difference in means as small as 0.2 standard deviations with $P(\text{CS}) \geq 0.99$. The table for $\mathcal{N}_{\text{B}}$ calls for $n = 361$ observations per population.

If, after taking $n = 361$ obns, we find that $\bar{Y}_1 = 13.2$, $\bar{Y}_2 = 9.8$, $\bar{Y}_3 = 16.1$, and $\bar{Y}_4 = 12.1$, then we select population 3 as the best.

Note that increasing $\delta^\star$ and/or decreasing $P^\star$ requires a smaller n. For example, when $\delta^\star/\sigma = 0.6$ and $P^\star = 0.95$, $\mathcal{N}_{\text{B}}$ requires only $n = 24$ observations per population. $\square$

Robustness of Procedure: How does $\mathcal{N}_{\text{B}}$ do under different types of violations of the underlying assumptions on which it's based?

- Lack of normality — not so bad.
- Different variances — sometimes a big problem.
- Dependent data — usually a nasty problem (e.g., in simulations).

Two-Stage Procedure $\mathcal{N}_{\mathrm{R}}$ (Rinott 1979)

Assumes populations have *unknown and unequal variances.* Takes a first stage of observations to estimate the variances of each system, and then uses those estimates to determine how many observations to take in the second stage — the higher the variance estimate, the more observations needed.

For the given k, specify $(P^\star, \delta^\star)$, and a common first-stage sample size $n_0 \geq 2$.

Look up the constant $g(P^\star, n_0, k)$ in an appropriate table or (if you have the urge) solve the following equation for g:

$$\int_0^\infty \int_0^\infty \Phi^{k-1}\left(\frac{g}{(n_0-1)(\frac{1}{x}+\frac{1}{y})}\right) f(x) f(y)\, dx\, dy \;=\; P^\star,$$

where $f(\cdot)$ is the $\chi^2(n_0-1)$ p.d.f.

Find the Normal Distribution with the Largest Mean

Take an i.i.d. sample $Y_{i1}, Y_{i2}, \ldots, Y_{in_0}$ from each of the k scenarios simulated independently.

Calculate the first-stage sample means and variances,

$$\bar{Y}_i(n_0) \;=\; \frac{1}{n_0}\sum_{j=1}^{n_0} Y_{ij} \quad \text{and} \quad S_i^2 \;=\; \frac{\sum_{j=1}^{n_0}\left(Y_{ij} - \bar{Y}_i(n_0)\right)^2}{n_0 - 1},$$

and then the final sample sizes

$$N_i \;=\; \max\left\{n_0, \left\lceil (gS_i/\delta^\star)^2 \right\rceil\right\}, \quad i = 1, 2, \ldots, k.$$

Take $N_i - n_0$ additional i.i.d. observations from scenario i, independently of the first-stage sample and the other scenarios, for $i = 1, 2, \ldots, k$.

Find the Normal Distribution with the Largest Mean

Compute overall sample means $\bar{\bar{Y}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}, \forall i$.

Select the scenario with the largest $\bar{\bar{Y}}_i$ as best.

Bonus: Simultaneously form MCP confidence intervals

$$\mu_i - \max_{j \neq i} \mu_j \in \left[-\left(\bar{\bar{Y}}_i - \max_{j \neq i} \bar{\bar{Y}}_j - \delta^\star \right)^- , \left(\bar{\bar{Y}}_i - \max_{j \neq i} \bar{\bar{Y}}_j + \delta^\star \right)^+ \right]$$

$\forall i$, where $(a)^+ \equiv \max\{0, a\}$ and $-(b)^- \equiv \min\{0, b\}$.

Find the Normal Distribution with the Largest Mean

$P^\star$	n_0	k					
		2	3	4	5	6	7
0.95	9	2.656	3.226	3.550	3.776	3.950	4.091
	10	2.614	3.166	3.476	3.693	3.859	3.993
	11	2.582	3.119	3.420	3.629	3.789	3.918
	12	2.556	3.082	3.376	3.579	3.734	3.860
	13	2.534	3.052	3.340	3.539	3.690	3.812
	14	2.517	3.027	3.310	3.505	3.654	3.773
	15	2.502	3.006	3.285	3.477	3.623	3.741
	16	2.489	2.988	3.264	3.453	3.597	3.713
	17	2.478	2.973	3.246	3.433	3.575	3.689
	18	2.468	2.959	3.230	3.415	3.556	3.669
	19	2.460	2.948	3.216	3.399	3.539	3.650
	20	2.452	2.937	3.203	3.385	3.523	3.634
	30	2.407	2.874	3.129	3.303	3.434	3.539
	40	2.386	2.845	3.094	3.264	3.392	3.495
	50	2.373	2.828	3.074	3.242	3.368	3.469

g Constant Required by $\mathcal{N}_{\mathrm{R}}$

Find the Normal Distribution with the Largest Mean

Example: A Simulation Study of Airline Reservation Systems

Consider $k = 4$ different airline reservation systems.

Objective: Find the system with the *largest* expected time to failure (E[TTF]). Let μ_i denote the E[TTF] for system i.

From past experience we know that the E[TTF]'s are roughly 100,000 minutes (about 70 days) for all four systems.

Goal: Select the best system with probability at least $P^\star = 0.90$ if the difference in the expected failure times for the best and second best systems is $\geq \delta^\star = 3000$ minutes (about two days).

The competing systems are sufficiently complicated that computer simulation is required to analyze their behavior.

Find the Normal Distribution with the Largest Mean

Let T_{ij} $(1 \leq i \leq 4, j \geq 1)$ denote the observed time to failure from the jth independent simulation replication of system i.

Application of the Rinott procedure $\mathcal{N}_{\mathrm{R}}$ requires i.i.d. normal observations from each system.

If each simulation replication is initialized from a particular system under the same operating conditions, but with independent random number seeds, the resulting $T_{i1}, T_{i2}, \ldots$ will be i.i.d. for each system.

However, the T_{ij} aren't normal — in fact, they're skewed right.

Find the Normal Distribution with the Largest Mean

Instead of using the raw T_{ij} in $\mathcal{N}_{\mathrm{R}}$, apply the procedure to the so-called *macroreplication* estimators of the μ_i.

These estimators group the $\{T_{ij} : j \geq 1\}$ into disjoint *batches* and use the batch averages as the "data" to which $\mathcal{N}_{\mathrm{R}}$ is applied.

Fix a number m of simulation replications that comprise each macroreplication (that is, m is the batch size) and let

$$Y_{ij} \equiv \frac{1}{m} \sum_{k=1}^{m} T_{i,(j-1)m+k}, \quad 1 \leq i \leq 4,\ 1 \leq j \leq b_i,$$

where b_i is the number of macroreplications to be taken from system i.

Find the Normal Distribution with the Largest Mean

The macroreplication estimators from the ith system, $Y_{i1}, Y_{i2}, \ldots, Y_{ib_i}$, are i.i.d. with expectation μ_i.

If m is sufficiently large, say at least 20, then the CLT yields approximate normality for each Y_{ij}.

No assumptions on the variances of the macroreplications.

To apply $\mathcal{N}_{\mathrm{R}}$, first conduct a pilot study to serve as the first stage of the procedure. Each system was run for $n_0 = 20$ macroreplications with each macroreplication consisting of the averages of $m = 20$ simulations of the system.

Rinott table with $k = 4$ and $P^\star = 0.90$ gives $g = 2.720$.

The total sample sizes N_i are computed for each system and are displayed in the summary table.

Find the Normal Distribution with the Largest Mean

i	1	2	3	4
$\bar{Y}_i(n_0)$	108286	107686	96167	89747
S_i	29157	24289	25319	20810
N_i	699	485	527	356
$\bar{\bar{Y}}_i$	110816	106411	99093	86568
std. error	872	1046	894	985

Summary of Airline Rez Example

E.g., System 2 requires an *additional* $N_2 - 20 = 465$ macroreplications in the second stage (each macroreplication again being the average of $m = 20$ system simulations).

In all, a total of about 40,000 simulations of the four systems were required to implement procedure $\mathcal{N}_{\mathrm{R}}$. The combined sample means for each system are listed in row 4 of the summary table.

Clearly establish System 1 as having the largest E[TTF]. □

Multi-Stage Procedure $\mathcal{N}_{\mathrm{KN}}$ (Kim & Nelson 2001)

Very efficient procedure. Takes observations from each population one-at-a-time, and eliminates populations that appear to be noncompetitive along the way.

Assumes populations have unknown (unequal) variances.

For the given k, specify $(P^\star, \delta^\star)$, and a common initial sample size from each scenario $n_0 \geq 2$.

To begin with, calculate the constant

$$\eta \equiv \frac{1}{2}\left[\left(\frac{2(1-P^\star)}{k-1}\right)^{-2/(n_0-1)} - 1\right].$$

Initialize $I = \{1, 2, \ldots, k\}$ and let $h^2 \equiv 2\eta(n_0 - 1)$.

Find the Normal Distribution with the Largest Mean

Take an initial random sample of $n_0 \geq 2$ observations Y_{ij} $(1 \leq j \leq n_0)$ from population i $(1 \leq i \leq k)$.

For population i, compute the sample mean based on the n_0 observations, $\bar{Y}_i(n_0) = \sum_{j=1}^{n_0} Y_{ij}/n_0$ $(1 \leq i \leq k)$.

For all $i \neq \ell$, compute the sample variance of the difference between populations i and ℓ,

$$S_{i\ell}^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} \left(Y_{ij} - Y_{\ell j} - [\bar{Y}_i(n_0) - \bar{Y}_\ell(n_0)] \right)^2 .$$

For all $i \neq \ell$, set $N_{i\ell} = \left\lfloor h^2 S_{i\ell}^2 / (\delta^\star)^2 \right\rfloor$ and then $N_i = \max_{\ell \neq i} N_{i\ell}$.

Find the Normal Distribution with the Largest Mean

If $n_0 > \max_i N_i$, stop and select the population with the largest sample mean $\bar{Y}_i(n_0)$ as one having the largest mean. Otherwise, set the sequential counter $r = n_0$ and go to the Screening phase of the procedure.

Screening: Set $I^{\text{old}} = I$ and re-set

$$\begin{aligned} I \;\; = \;\; & \{i : i \in I^{\text{old}} \text{ and } \bar{Y}_i(r) \geq \bar{Y}_\ell(r) - W_{i\ell}(r), \\ & \qquad \text{for all } \ell \in I^{\text{old}}, \ell \neq i\}, \end{aligned}$$

where

$$W_{i\ell}(r) \;\; = \;\; \max\left\{0, \frac{\delta^\star}{2r}\left(\frac{h^2 S_{i\ell}^2}{(\delta^\star)^2} - r\right)\right\}.$$

Keep those surviving populations that aren't "too far" from the current leader.

Stopping Rule: If $|I| = 1$, then stop and select the treatment with index in I as having the largest mean.

If $|I| > 1$, take one additional observation $Y_{i,r+1}$ from each treatment $i \in I$.

Increment $r = r + 1$ and go to the screening stage if $r < \max_i N_i + 1$.

If $r = \max_i N_i + 1$, then stop and select the treatment associated with the largest $\bar{Y}_i(r)$ having index $i \in I$.

Normal Extensions

Correlation *between* populations.

Better fully sequential procedures.

Better elimination of populations that aren't competitive.

Different variance estimators.

Find the Bernoulli with the Largest Success Probability

Examples:

- Which anti-cancer drug is most effective?
- Which simulated system is most likely to meet design specs?

There are 100's of such procedures. Highlights:

- Single-Stage Procedure (Sobel and Huyett 1957)
- Sequential Procedure (Bechhofer, Kiefer, Sobel 1968)
- "Optimal" Procedures (Bechhofer et al., 1980's)

Again use the *indifference-zone* approach.

Find the Bernoulli with the Largest Success Probability

We have k competing Bern populations with success parameters $p_1, p_2, \ldots, p_k$. Denote the ordered p's by $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[k]}$.

Goal: Select the population having the largest probability $p_{[k]}$.

Probability Requirement: For specified constants $(P^\star, \Delta^\star)$ with $1/k < P^\star < 1$ and $0 < \Delta^\star < 1$, we require

$$P(\text{CS}) \;\geq\; P^\star \quad \text{whenever} \quad p_{[k]} - p_{[k-1]} \;\geq\; \Delta^\star.$$

The prob req't is defined in terms of the difference $p_{[k]} - p_{[k-1]}$, and we interpret $\Delta^\star$ as the "smallest difference worth detecting."

A Single-Stage Procedure $\mathcal{B}_{\mathrm{SH}}$ (Sobel and Huyett 1957)

For the specified $(P^\star, \Delta^\star)$, find n from a table.

Take a sample of n observations X_{ij} $(1 \leq j \leq n)$ in a *single* stage from each population $(1 \leq i \leq k)$.

Calculate the k sample sums $Y_{in} = \sum_{j=1}^{n} X_{ij}$.

Select the treatment that yielded the largest Y_{in} as the one associated with $p_{[k]}$; in the case of ties, randomize.

Find the Bernoulli with the Largest Success Probability

k	$\Delta^\star$	$P^\star$						
		0.60	0.75	0.80	0.85	0.90	0.95	0.99
3	0.10	20	52	69	91	125	184	327
	0.20	5	13	17	23	31	46	81
	0.30	3	6	8	10	14	20	35
	0.40	2	4	5	6	8	11	20
	0.50	2	3	3	4	5	7	12
4	0.10	34	71	90	114	150	212	360
	0.20	9	18	23	29	38	53	89
	0.30	4	8	10	13	17	23	39
	0.40	3	5	6	7	9	13	21
	0.50	2	3	4	5	6	8	13

Smallest n for $\mathcal{B}_{\mathrm{SH}}$ to Guarantee Probability Requirement

Example: Suppose we want to select the best of $k = 4$ treatments with probability at least $P^\star = 0.95$ whenever $p_{[4]} - p_{[3]} \geq 0.10$.

The table shows that we need $n = 212$ observations.

Suppose that, at the end of sampling, we have $Y_{1,212} = 70$, $Y_{2,212} = 145$, $Y_{3,212} = 95$, and $Y_{4,212} = 102$.

Then we select population 2 as the best. $\square$

A Curtailment Trick (Bechhofer and Kulkarni)

Idea: Do the single-stage procedure, except stop sampling when the guy in second place can *at best tie*.

This is called *curtailment* — you might as well stop because it won't be possible for the outcome to change (except if there's a tie, which doesn't end up mattering).

Turns out curtailment gives the same $P(\text{CS})$ as the single-stage procedure, but a lower expected number of observations ($\leq n$).

Example (cont'd): Recall that for $k = 4$, $P^\star = 0.95$, and $\Delta^\star = 0.10$, the single-sample procedure required us to take $n = 212$ observations.

Suppose that, at the end of just 180 samples from each population, we have the intermediate result $Y_{1,180} = 50$, $Y_{2,180} = 130$, $Y_{3,180} = 74$, and $Y_{4,180} = 97$.

We *stop sampling right now* and select population 2 as the best because it's not possible for population 4 to catch up in the remaining $212 - 180 = 32$ observations — big savings! $\square$

A Sequential Procedure $\mathcal{B}_{\mathrm{BKS}}$ (BKS 1968)

New Prob Requirement: For specified $(P^\star, \theta^\star)$ with $1/k < P^\star < 1$ and $\theta^\star > 1$, we require $P(\mathrm{CS}) \geq P^\star$ whenever the *odds ratio*

$$\frac{p_{[k]}/(1-p_{[k]})}{p_{[k-1]}/(1-p_{[k-1]})} \geq \theta^\star.$$

The procedure proceeds in stages, where we take one Bernoulli observation from each of the populations.

It's even more efficient than curtailment!

Find the Bernoulli with the Largest Success Probability

At the mth stage of experimentation ($m \geq 1$),

- Observe the random Bernoulli vector $(X_{1m}, \ldots, X_{km})$.
- Compute the sums $Y_{im} = \sum_{j=1}^{m} X_{ij}$ $(1 \leq i \leq k)$, and denote the ordered sums by $Y_{[1]m} \leq \cdots \leq Y_{[k]m}$.
- Stop if

$$Z_m \equiv \sum_{i=1}^{k-1} (1/\theta^\star)^{Y_{[k]m} - Y_{[i]m}} \leq \frac{1 - P^\star}{P^\star}.$$

Let N be the (random) stage m when the procedure stops.

Select the population yielding $Y_{[k]N}$ as the one associated with $p_{[k]}$.

Example: For $k = 3$ and $(P^\star, \theta^\star) = (0.75, 2)$, suppose the following sequence of vector-observations is obtained using $\mathcal{B}_{\mathrm{BKS}}$.

m	X_{1m}	X_{2m}	X_{3m}	Y_{1m}	Y_{2m}	Y_{3m}	Z_m
1	1	0	1	1	0	1	1.5
2	0	1	1	1	1	2	1.0
3	0	1	1	1	2	3	0.75
4	0	0	1	1	2	4	0.375
5	1	1	1	2	3	5	0.375
6	1	0	1	3	3	6	0.25

Since $Z_6 \leq (1 - P^\star)/P^\star = 1/3$, sampling stops at stage $N = 6$ and population 3 is selected as best. $\square$

Bernoulli Extensions

Correlation *between* populations.

More-efficient sequential procedures.

Elimination of populations that aren't competitive.

Find the Most Probable Multinomial Cell

Examples:

- Who is the most popular political candidate?
- Which television show is most watched during a particular time slot?
- Which simulated warehouse configuration is most likely to maximize throughput?

Yet again, use the indifference-zone approach.

Experimental Set-Up:

- k possible outcomes (categories).
- p_i is the probability of the ith category.
- n independent replications of the experiment.
- Y_i is the number of outcomes falling in category i after the n observations have been taken.

Definition: If the k-variate discrete vector random variable $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_k)$ has the probability mass function

$$P\{Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k\} = \frac{n!}{\prod_{i=1}^{k} y_i!} \prod_{i=1}^{k} p_i^{y_i},$$

then $\boldsymbol{Y}$ has a *multinomial* distribution with parameters n and $\boldsymbol{p} = (p_1, \ldots, p_k)$, where $\sum_{i=1}^{k} p_i = 1$ and $p_i > 0$ for all i.

Find the Most Probable Multinomial Cell

Example: Suppose three of the faces of a fair die are red, two are blue, and one is violet, i.e., $\boldsymbol{p} = (3/6, 2/6, 1/6)$.

Toss it $n = 5$ times. Then the probability of observing exactly three reds, no blues and two violets is

$$P\{\boldsymbol{Y} = (3, 0, 2)\} = \frac{5!}{3!0!2!}(3/6)^3(2/6)^0(1/6)^2 = 0.03472. \quad \square$$

Example (continued): Suppose we did not know the probabilities for red, blue, and violet in the previous example and that we want to select the most probable color.

The selection rule is to choose the color that occurs the most frequently during the five trials, using randomization to break ties.

Find the Most Probable Multinomial Cell

Let $\boldsymbol{Y} = (Y_r, Y_b, Y_v)$ denote the number of occurrences of (red, blue, violet) in five trials. The probability that we correctly select red is...

$$
\begin{aligned}
& P\{\text{red wins in 5 trials}\} \\
= \;& P\{Y_r > Y_b \text{ and } Y_v\} + 0.5P\{Y_r = Y_b, Y_r > Y_v\} \\
& \quad + 0.5P\{Y_r > Y_b, Y_r = Y_v\} \\
= \;& P\{\boldsymbol{Y} = (5,0,0), (4,1,0), (4,0,1), (3,2,0), (3,1,1), (3,0,2)\} \\
& \quad + 0.5P\{\boldsymbol{Y} = (2,2,1)\} + 0.5P\{\boldsymbol{Y} = (2,1,2)\}.
\end{aligned}
$$

We can list the outcomes favorable to a *correct selection* (CS) of red, along with the associated probabilities, randomizing in case of ties...

Find the Most Probable Multinomial Cell

Outcome (red, blue, violet)	Contribution to $P\{$red wins in 5 trials$\}$
(5,0,0)	0.03125
(4,1,0)	0.10417
(4,0,1)	0.05208
(3,2,0)	0.13889
(3,1,1)	0.13889
(3,0,2)	0.03472
(2,2,1)	(0.5)(0.13889)
(2,1,2)	(0.5)(0.06944)
	0.60416

The probability of correctly selecting red as the most probable color based on $n = 5$ trials is 0.6042. This $P(\text{CS})$ can be increased by increasing the sample size n. $\square$

Find the Most Probable Multinomial Cell

Example: The most probable alternative might be preferable to that having the largest expected value.

Consider two inventory policies, A and B, where

$$\text{Profit from } A \quad = \quad \$5 \text{ with probability } 1$$

$$\text{Profit from } B \quad = \quad \begin{cases} \$0 & \text{with probability } 0.99 \\ \$1000 & \text{with probability } 0.01. \end{cases}$$

Then

$$\text{E[Profit from } A] = \$5 \; < \; \text{E[Profit from } B] = \$10$$

$$P\{\text{Profit from } A > \text{Profit from } B\} \; = \; 0.99.$$

So $\text{E}[A] < \text{E}[B]$, but A wins almost all of the time. $\square$

Assumptions and Notation for Multinomial Selection

- $\boldsymbol{X}_j = (X_{1j}, \ldots, X_{kj})$ $(j \geq 1)$ are independent observations taken from a multinomial distribution having $k \geq 2$ categories with associated unknown probabilities $\boldsymbol{p} = (p_1, \ldots, p_k)$.
- $X_{ij} = 1$ [0] if category i does [does not] occur on the jth observation.
- The (unknown) ordered p_i's are $p_{[1]} \leq \cdots \leq p_{[k]}$.
- The category with $p_{[k]}$ is the *most probable* or *best*.
- The cumulative sum for category i after m multinomial observations have been taken is $Y_{im} = \sum_{j=1}^{m} X_{ij}$.
- The ordered Y_{im}'s are $Y_{[1]m} \leq \cdots \leq Y_{[k]m}$.

Indifference-Zone Procedures

Goal: Select the category associated with $p_{[k]}$.

A *correct selection* (CS) is made if the Goal is achieved.

Probability Requirement: For specified $(P^\star, \theta^\star)$ with $1/k < P^\star < 1$ and $\theta^\star > 1$, we require

$$P(\text{CS}) \geq P^\star \quad \text{whenever} \quad p_{[k]}/p_{[k-1]} \geq \theta^\star. \tag{3}$$

The probability in (3) depends on the entire vector $\boldsymbol{p}$ and on the number n of independent multinomial observations to be taken.

$\theta^\star$ is the "smallest $p_{[k]}/p_{[k-1]}$ ratio worth detecting."

Can consider various procedures to guarantee prob req't (3).

Single-Stage Procedure $\mathcal{M}_{\mathrm{BEM}}$

For the given k, $P^\star$ and $\theta^\star$, find n from the table (sort of from Bechhofer, Elmaghraby, and Morse 1959).

Take n multinomial observations $\boldsymbol{X}_j = (X_{1j}, \ldots, X_{kj})$ $(1 \le j \le n)$ in a *single* stage.

Calculate the ordered sample sums $Y_{[1]n} \le \cdots \le Y_{[k]n}$. Select the category with the largest sum, $Y_{[k]n}$, as the one associated with $p_{[k]}$, randomizing to break ties.

Remark: The n-values are computed so that $\mathcal{M}_{\mathrm{BEM}}$ achieves $P(\mathrm{CS}) \ge P^\star$ when the cell probabilities $\boldsymbol{p}$ are in the *least-favorable* (LF) configuration (Kesten and Morse 1959),

$$p_{[1]} = p_{[k-1]} = 1/(\theta^\star + k - 1) \text{ and } p_{[k]} = \theta^\star/(\theta^\star + k - 1). \quad (4)$$

Example: A soft drink producer wants to find the most popular of $k = 3$ proposed cola formulations.

The company will give a taste test to n people.

The sample size n is to be chosen so that $P(\text{CS}) \geq 0.95$ whenever the ratio of the largest to second largest true (but unknown) proportions is at least 1.4.

Entering the table with $k = 3$, $P^\star = 0.95$, and $\theta^\star = 1.4$, we find that $n = 186$ individuals must be interviewed.

If we find that $Y_{1,186} = 53$, $Y_{2,186} = 110$, and $Y_{3,186} = 26$, then we select formulation 2 as the best. $\square$

Find the Most Probable Multinomial Cell

$P^\star$	$\theta^\star$	$k=2$		$k=3$		$k=4$		$k=5$	
		n	n_0	n	n_0	n	n_0	n	n_0
	2.0	5	5	12	13	20	24	29	34
	1.8	5	7	17	18	29	35	41	50
0.75	1.6	9	9	26	32	46	57	68	86
	1.4	17	19	52	71	92	124	137	184
	1.2	55	67	181	285	326	495	486	730
	2.0	15	15	29	34	43	53	58	71
	1.8	19	27	40	50	61	75	83	104
0.90	1.6	31	41	64	83	98	126	134	172
	1.4	59	79	126	170	196	274	271	374
	1.2	199	267	437	670	692	1050	964	1460
	2.0	23	27	42	52	61	74	81	98
	1.8	33	35	59	71	87	106	115	142
0.95	1.6	49	59	94	125	139	180	185	240
	1.4	97	151	186	266	278	380	374	510
	1.2	327	455	645	960	979	1500	1331	2000

Sample Sizes n for $\mathcal{M}_{\mathrm{BEM}}$ and Truncation Numbers n_0 for $\mathcal{M}_{\mathrm{BG}}$ to Guarantee (3)

A Curtailed Procedure $\mathcal{M}_{\mathrm{BK}}$ (Bechhofer and Kulkarni 1984)

For the given k, specify n prior to the start of sampling.

At the mth stage of experimentation ($m \geq 1$), take the random observation $\boldsymbol{X}_m = (X_{1m}, \ldots, X_{km})$.

Calculate the sample sums Y_{im} through stage m ($1 \leq i \leq k$). Stop sampling at the first stage m for which there exists a category satisfying

$$Y_{im} \geq Y_{jm} + n - m \quad \text{for all } j \neq i \ (1 \leq i, j \leq k). \tag{5}$$

Let N (a random variable) denote the value of m at the termination of sampling. Select the category having the largest sum as the one associated with $p_{[k]}$, randomizing to break ties.

Remark: The LHS of (5) is the current total number of occurrences of category i; the RHS is the current total of category j *plus* the additional number of *potential* occurrences of j if all of the $(n - m)$ remaining outcomes after stage m were also to be associated with j.

Thus, *curtailment* takes place when one of the categories has sufficiently more successes than all of the other categories, i.e., sampling stops when the leader can do no worse than *tie*.

Procedure $\mathcal{M}_{\mathrm{BK}}$ saves observations and achieves the same $P(\mathrm{CS})$ as does $\mathcal{M}_{\mathrm{BEM}}$ with the same n. In fact,...

$$P\{\text{CS using } \mathcal{M}_{\mathrm{BK}} \,|\, \boldsymbol{p}\} \;=\; P\{\text{CS using } \mathcal{M}_{\mathrm{BEM}} \,|\, \boldsymbol{p}\}$$

and

$$\mathrm{E}\{N \text{ using } \mathcal{M}_{\mathrm{BK}} \,|\, \boldsymbol{p}\} \;\leq\; n \text{ using } \mathcal{M}_{\mathrm{BEM}} \,.$$

Example: For $k = 3$ and $n = 2$, stop sampling if

m	X_{1m}	X_{2m}	X_{3m}	Y_{1m}	Y_{2m}	Y_{3m}
1	1	0	0	1	0	0

and select category 1 because
$Y_{1m} = 1 \geq Y_{jm} + n - m = 0 + 2 - 1 = 1$ for $j = 2$ and 3. □

Example: For $k = 3$ and $n = 3$ or 4, stop sampling if

m	X_{1m}	X_{2m}	X_{3m}	Y_{1m}	Y_{2m}	Y_{3m}
1	0	1	0	0	1	0
2	0	1	0	0	2	0

and select category 2 because $Y_{2m} = 2 \geq Y_{jm} + n - m = 0 + n - 2$ for $n = 3$ or $n = 4$ and both $j = 1$ and 3. □

Example: For $k = 3$ and $n = 3$ suppose that

m	X_{1m}	X_{2m}	X_{3m}	Y_{1m}	Y_{2m}	Y_{3m}
1	1	0	0	1	0	0
2	0	0	1	1	0	1
3	0	1	0	1	1	1

Because $Y_{13} = Y_{23} = Y_{33} = 1$, we stop sampling and randomize among the three categories. $\square$

Sequential Procedure with Curtailment $\mathcal{M}_{\mathrm{BG}}$ (Bechhofer and Goldsman 1986)

For the given k and specified $(P^\star, \theta^\star)$, find the *truncation number* n_0 from the table.

At the mth stage of experimentation ($m \geq 1$), take the random observation $\boldsymbol{X}_m = (X_{1m}, \ldots, X_{km})$.

Calculate the ordered category totals $Y_{[1]m} \leq \cdots \leq Y_{[k]m}$ and

$$Z_m \;=\; \sum_{i=1}^{k-1} (1/\theta^\star)^{(Y_{[k]m} - Y_{[i]m})}.$$

Stop sampling at the first stage when *either*

$$Z_m \;\leq\; (1 - P^\star)/P^\star \quad \text{or} \quad Y_{[k]m} - Y_{[k-1]m} \;\geq\; n_0 - m. \qquad (6)$$

Let N denote the value of m at the termination of sampling. Select the category that yielded $Y_{[k]N}$ as the one associated with $p_{[k]}$; randomize in the case of ties.

Remark: The truncation numbers n_0 given in the previous table are calculated assuming that Procedure $\mathcal{M}_{\mathrm{BG}}$ has the same LF-configuration (3) as does $\mathcal{M}_{\mathrm{BEM}}$. (This hasn't been proven yet.)

Example: Suppose $k = 3$, $P^\star = 0.75$, and $\theta^\star = 3.0$. The table tells us to truncate sampling at $n_0 = 5$ observations. For the data

m	X_{1m}	X_{2m}	X_{3m}	Y_{1m}	Y_{2m}	Y_{3m}
1	0	1	0	0	1	0
2	0	1	0	0	2	0

we stop sampling by the first criterion in (6) because $Z_2 = (1/3)^2 + (1/3)^2 = 2/9 \leq (1 - P^\star)/P^\star = 1/3$, and we select category 2. $\square$

Example: Again suppose $k = 3$, $P^\star = 0.75$, and $\theta^\star = 3.0$ (so that $n_0 = 5$). For the data

m	X_{1m}	X_{2m}	X_{3m}	Y_{1m}	Y_{2m}	Y_{3m}
1	0	1	0	0	1	0
2	1	0	0	1	1	0
3	0	1	0	1	2	0
4	1	0	0	2	2	0
5	1	0	0	3	2	0

we stop sampling by the second criterion in (6) because $m = n_0 = 5$ observations, and we select category 1. $\square$

Example: Yet again suppose $k = 3$, $P^\star = 0.75$, and $\theta^\star = 3.0$ (so that $n_0 = 5$). For the data

m	X_{1m}	X_{2m}	X_{3m}	Y_{1m}	Y_{2m}	Y_{3m}
1	0	1	0	0	1	0
2	1	0	0	1	1	0
3	0	1	0	1	2	0
4	1	0	0	2	2	0
5	0	0	1	2	2	1

we stop according to the second criterion in (6) because $m = n_0 = 5$. However, we now have a tie between $Y_{1,5}$ and $Y_{2,5}$ and thus randomly select between categories 1 and 2. $\square$

Example: Still yet again suppose $k = 3$, $P^\star = 0.75$, and $\theta^\star = 3.0$ (so that $n_0 = 5$). Suppose we observe

m	X_{1m}	X_{2m}	X_{3m}	Y_{1m}	Y_{2m}	Y_{3m}
1	0	1	0	0	1	0
2	1	0	0	1	1	0
3	0	1	0	1	2	0
4	0	0	1	1	2	1

Because categories 1 and 3 can do no better than tie category 2 (if we were to take the potential remaining $n_0 - m = 5 - 4 = 1$ observation), the second criterion in (6) tells us to stop; we select category 2. $\square$

Remark: Procedure $\mathcal{M}_{\mathrm{BG}}$ usually requires fewer observations than $\mathcal{M}_{\mathrm{BEM}}$.

Example: Suppose $k = 4$, $P^\star = 0.75$, $\theta^\star = 1.6$.

The single-stage procedure $\mathcal{M}_{\mathrm{BEM}}$ requires 46 observations to guarantee (3).

Procedure $\mathcal{M}_{\mathrm{BG}}$ (with a truncation number of $n_0 = 57$) has $\mathrm{E}[N|\mathrm{LF}] = 31.1$ and $\mathrm{E}[N|\mathrm{EP}] = 37.7$ for $\boldsymbol{p}$ in the LF configuration (4) and *equal-probability* (EP) configuration, $p_{[1]} = p_{[k]}$, respectively.

Applications

Let's take i.i.d. vector-observations $\boldsymbol{W}_j = (W_{1j}, \ldots, W_{kj})$ $(j \geq 1)$, where the W_{ij} can be either discrete or continuous.

For a particular vector-observation $\boldsymbol{W}_j$, suppose the experimenter can determine which of the k observations W_{ij} $(1 \leq i \leq k)$ is the "most desirable." The term "most desirable" is based on some criterion of goodness designated by the experimenter, and it can be quite general, e.g.,. . .

- The largest crop yield based on a vector-observation of k agricultural plots using competing fertilizers.
- The smallest sample average customer waiting time based on a simulation run of each of k competing queueing strategies.
- The smallest estimated variance of customer waiting times (from the above simulations).

Find the Most Probable Multinomial Cell

For a particular vector-observation $\boldsymbol{W}_j$, suppose $X_{ij} = 1$ or 0 according as W_{ij} $(1 \leq i \leq k)$ is the "most desirable" of the components of $\boldsymbol{W}_j$ or not. Then $(X_{1j}, \ldots, X_{kj})$ $(j \geq 1)$ has a multinomial distribution with probability vector $\boldsymbol{p}$, where

$$p_i \;=\; P\{W_{i1} \text{ is the "most desirable" component of } \boldsymbol{W}_1\}.$$

Selection of the category corresponding to the largest p_i can be thought of as that of finding the component having the highest probability of yielding the "most desirable" observation of those from a particular vector-observation. This problem can be approached using the multinomial selection methods described in this module.

Find the Most Probable Multinomial Cell

Example: Suppose we want to find which of $k = 3$ *job shop* configurations is most likely to give the shortest expected times-in-system for a certain manufactured product. Because of the complicated configurations of the candidate job shops, it is necessary to simulate the three competitors. Suppose the jth simulation run of configuration i yields W_{ij} ($1 \leq i \leq 3$, $j \geq 1$), the proportion of 1000 times-in-system greater than 20 minutes.

Management has decided that the "most desirable" component of $\boldsymbol{W}_j = (W_{1j}, W_{2j}, W_{3j})$ will be that component corresponding to $\min_{1 \leq i \leq 3} W_{ij}$.

Find the Most Probable Multinomial Cell

If p_i denotes the probability that configuration i yields the smallest component of $\boldsymbol{W}_j$, then we seek to select the configuration corresponding to $p_{[3]}$. Specify $P^\star = 0.75$ and $\theta^\star = 3.0$. The truncation number from the table for $\mathcal{M}_{\mathrm{BG}}$ is $n_0 = 5$. We apply the procedure to the data

m	W_{1m}	W_{2m}	W_{3m}	X_{1m}	X_{2m}	X_{3m}	Y_{1m}	Y_{2m}	Y_{3m}
1	0.13	0.09	0.14	0	1	0	0	1	0
2	0.24	0.10	0.07	0	0	1	0	1	1
3	0.17	0.11	0.12	0	1	0	0	2	1
4	0.13	0.08	0.02	0	0	1	0	2	2
5	0.14	0.13	0.15	0	1	0	0	3	2

. . . and select shop configuration 2. □