Selecting the Best Simulated System

Dave Goldsman[†]

Georgia Tech

July 3–4, 2003

[†]Barry Nelson

Outline of Talk

1. Introduction. Look at procedures to select the...

2. Normal Population with the Largest Mean

3. Binomial Popn with the Largest Success Parameter

4. Most Probable Multinomial Cell.

5. Simulation Applications.

Introduction

Statistics / Simulation experiments are typically performed to compare a "small" number of system designs — say 2 to 200.

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

We present **ranking & selection** and **multiple comparison** procedures.

Examples:

Which of 10 fertilizers produces the largest mean crop yield? (Normal)

Find the pain reliever that has the highest probability of giving relief. (Binomial)

Which candidate is the most popular? (Multinomial)

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

* Guarantee a probability of a correct selection.

MCPs treat the comparison problem as an inference problem.

* Account for simultaneous errors.

R&S and MCPs are both relevant in simulation:

* Normally distributed data by batching.

* Independence by controlling random numbers.

* Multiple-stage sampling by retaining the seeds.

2. Find the Normal Distrn with the Largest Mean

We present procedures for selecting the *normal* distribution that has the largest mean.

We use the *indifference-zone* approach.

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

Notation: Denote the vector of means by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_t)$ and the vector of variances by $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_t^2)$. The ordered μ_i's are

$$\mu_{[1]} \leq \cdots \leq \mu_{[t]}.$$

The treatment having mean $\mu_{[t]}$ is the "best" treatment.

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

Definition: A *correct selection* (CS) is said to be made if the Goal is achieved.

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

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

The probability in (1) depends on the differences $\mu_i - \mu_j$ ($i \neq j$, $1 \leq i, j \leq t$), the sample size n, and $\boldsymbol{\sigma}^2$. The constant $\delta^\star$ can be thought of as the "smallest difference worth detecting."

Parameter configurations $\boldsymbol{\mu}$ satisfying $\mu_{[t]} - \mu_{[t-1]} \geq \delta^\star$ are in the *preference-zone* for a correct selection; configurations satisfying $\mu_{[t]} - \mu_{[t-1]} < \delta^\star$ are in the *indifference-zone*. Any procedure that guarantees (1) is said to be employing the *indifference-zone* approach.

There are $>$ 100 such procedures. Highlights:

* Single-Stage Procedure (Bechhofer 1954)

* Two-Stage Procedure (Rinott 1979)

* Sequential Procedure (Nelson and friends, 2001)

Single-Stage Procedure $\mathcal{N}_B$ (Bechhofer 1954)

Assumes popns have *common known variance*.

For the given t and specified $(\delta^\star/\sigma, P^\star)$, determine sample size n (usually from a table).

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

Calculate the t sample means

$$\overline{Y}_i = \sum_{j=1}^{n} Y_{ij}/n \quad (1 \le i \le t).$$

Select the popn that yielded the largest sample mean, $\overline{Y}_{[t]} = \max\{\overline{Y}_1, \ldots, \overline{Y}_t\}$, as the one associated with $\mu_{[t]}$.

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

t	$P^\star$	$\delta^\star/\sigma$									
		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 Popn Required by $\mathcal{N}_B$

Remark: Don't really need the above table. . .

$$n \;=\; \lceil 2(\sigma Z^{(1-P^\star)}_{t-1,1/2}/\delta^\star)^2 \rceil$$

where $Z^{(1-P^\star)}_{t-1,1/2}$ is a special case of the upper equicoordinate point of a certain multivariate normal distribution.

The constant $Z_{p,\rho}^{(\alpha)}$ is determined to satisfy the probability requirement (1) for any true configuration of means satisfying

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

Configurations (2) are termed *least-favorable* (LF) because, for fixed n, they minimize the $P\{\text{CS}\}$ among all configurations satisfying the preference-zone requirement $\mu_{[t]} - \mu_{[t-1]} \geq \delta^\star$.

Example: Suppose $t = 4$ and we want to detect a difference in means as small as 0.2 standard deviations with probability 0.99. The table shows that $\mathcal{N}_B$ calls for $n = 361$ observations per popn. Increasing $\delta^\star$ and/or decreasing $P^\star$ requires a smaller n. For example, when $\delta^\star/\sigma = 0.6$ and $P^\star = 0.95$, $\mathcal{N}_B$ requires only $n = 24$ observations per treatment.

Robustness of Normal Theory Procedure

How does $\mathcal{N}_B$ 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 very big problem (especially in simulations).

Two-Stage Procedure $\mathcal{N}_R$ (Rinott 1979)

Assumes popns have unknown (unequal) variances.

For the given t, specify $(\delta^\star, P^\star)$.

Specify a common first-stage sample size $n_0 \geq 2$.

Look up the constant $h(P^\star, n_0, t)$ in an appropriate table.

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

Calculate the first-stage sample means

$$\bar{Y}_i^{(1)} = \sum_{j=1}^{n_0} Y_{ij}/n_0,$$

and marginal sample variances

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

for $i = 1, 2, \ldots, t$.

Compute the final sample sizes

$$N_i \;=\; \max\left\{n_0, \left\lceil (hS_i/\delta^\star)^2 \right\rceil\right\}$$

for $i = 1, 2, \ldots, t$, where $\lceil \cdot \rceil$ is the integer "round-up" function.

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

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

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

Bonus: Simultaneously form "multiple comparisons with the best" confidence intervals

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

for $i = 1, 2, \ldots, t$, where $(a)^+ = \max\{0, a\}$ and $-(b)^- = \min\{0, b\}$.

$P^\star$	$n_0 = \nu + 1$	t = 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

h Constant Required by $\mathcal{N}_R$

Example: A Simulation Study of Airline Reservation Systems

Consider $t = 4$ 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 $(1 \leq i \leq 4)$.

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.

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 $\mathcal{N}_R$ 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.

Instead of using the raw T_{ij} in procedure $\mathcal{N}_R$, 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 $\mathcal{N}_R$ is applied.

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

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

$(1 \leq i \leq 4,\ 1 \leq j \leq b_i)$ where b_i is the number of macroreplications to be taken from system i.

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 are made concerning the variances of the macroreplications.

To apply $\mathcal{N}_R$, 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 $t = 4$ and $P^\star = 0.90$ gives $h = 2.720$.

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

i	1	2	3	4
$\overline{y}_i^{(1)}$	108286.	107686.	96167.7	89747.9
s_i	29157.3	24289.9	25319.5	20810.8
N_i	699	485	527	356
$\overline{y}_i$	110816.5	106411.8	99093.1	86568.9
$s_{\overline{y}_i}$	872.0	1046.5	894.2	985.8

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 $\mathcal{N}_R$. 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 $\mathcal{N}_{KN}$ (Kim & Nelson 2001)

Assumes popns have unknown (unequal) variances.

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

Calculate the constant

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

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

Take a random sample of n_0 observations Y_{ij} $(1 \leq j \leq n_0)$ from population i $(1 \leq i \leq t)$.

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

For all $i \neq \ell$, compute the sample variance of the difference between treatments 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

$$N_{i\ell} \;=\; \left\lfloor h^2 S_{i\ell}^2 / \delta^{\star 2} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ is the floor function, and

$$N_i \;=\; \max_{\ell \neq i} N_{i\ell}.$$

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

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

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 popns that aren't competitive.

Different variance estimators.

3. Find the Bernoulli Distrn with the Largest Success Probability

Again use the *indifference-zone* approach.

Examples:

* Which anti-cancer drug is most effective?

* Which simulated system is most likely to meet certain design specs?

There are $>$ 100 such procedures. Highlights:

* Single-Stage Procedure (Sobel and Huyett, 1957)

* Sequential Procedure (Bechhofer, Kiefer, Sobel, 1968)

* "Optimal" Procedures (Bechhofer, et al., 1980's)

A Single-Stage Procedure (Sobel and Huyett 1957)

t competing Bern populations Π_i with success parameters $p_1, p_2, \ldots, p_t$.

Ordered p's: $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[t]}$.

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

$$P\{\text{CS}\} \ \geq \ P^\star \text{ whenever } p_{[t]} - p_{[t-1]} \ \geq \ \Delta^\star. \qquad (3)$$

Note that the indifference requirement is defined in terms of the difference $p_{[t]} - p_{[t-1]}$.

The probability in (3) depends on the entire vector $\boldsymbol{p} = (p_1, p_2, \ldots, p_t)$ and on the common number n of independent observations taken from each of the t treatments.

The constant $\Delta^\star$ can be interpreted as the "smallest $p_{[t]} - p_{[t-1]}$ difference worth detecting."

Procedure $\mathcal{B}_{SH}$

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

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

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

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

t	$\Delta^\star$	$P^\star$						
		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 Sample Size for $\mathcal{B}_{SH}$ to Guarantee PR

Example: Suppose we want to select the best of $t = 4$ 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 Π_2 as the best.

A Sequential Procedure (BKS 1968).

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

$$P\{\text{CS}\} \;\geq\; P^\star \tag{4}$$

whenever the *odds ratio*

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

Procedure $\mathcal{B}_{BKS}$

For the given t, specify $(\theta^\star, P^\star)$.

At the mth stage of experimentation ($m \geq 1$), observe the random Bernoulli vector $(X_{1m}, \ldots, X_{tm})$.

Let $Y_{im} = \sum_{j=1}^{m} X_{ij}$ $(1 \leq i \leq t)$ and denote the ordered Y_{im}-values by $Y_{[1]m} \leq \cdots \leq Y_{[t]m}$.

After the mth stage of experimentation, compute

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

Stop at the first value of m (call it N) for which $Z_m \le (1 - P^\star)/P^\star$. Note that N is a random variable.

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

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

m	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}	$y_{[3]m} - y_{[2]m}$	$y_{[3]m} - y_{[1]m}$	z_m
1	1	0	1	1	0	1	0	1	1.5
2	0	1	1	1	1	2	1	1	1.0
3	0	1	1	1	2	3	1	2	0.75
4	0	0	1	1	2	4	2	3	0.375
5	1	1	1	2	3	5	2	3	0.375
6	1	0	1	3	3	6	3	3	0.25

Since $z_6 \le (1 - P^\star)/P^\star = 1/3$, sampling stops at stage $N = 6$ and treatment Π_3 is selected as best.

Bernoulli Extensions

Correlation *between* populations.

More-efficient sequential procedures.

Elimination of populations that aren't competitive.

Multinomial Selection and Screening

Overview

We present procedures for selecting the *multinomial* category that has the largest probability of occurrence. We use both the indifference-zone and subset selection approaches.

Example:

- Who is the most popular political candidate?

- Which television show is most watched during a particular time slot?

- Which simulated warehouse configuration maximizes throughput?

Organization

1. Motivational examples.
2. Notation.
3. Several procedures.
4. Simulation applications.

Motivational Examples

Experimental Set-Up:

- t 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.

The t-variate discrete vector random variable $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_t)$ has the probability mass function

$$P\{Y_1 = y_1, Y_2 = y_2, \ldots, Y_t = y_t\} = \frac{n!}{\prod_{i=1}^{t} y_i!} \prod_{i=1}^{t} p_i^{y_i},$$

and $\boldsymbol{Y}$ has a *multinomial* distribution with parameters n and $\boldsymbol{p} = (p_1, \ldots, p_t)$.

Example: Suppose three of the faces of a fair die are red, two are blue, and one is green, 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 greens is

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

Example (continued): Suppose we did not know the probabilities for red, blue and green 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.

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

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

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

Outcome (red, blue, green)	Contribution to $P\{\text{red wins in 5 trials}\}$
(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.

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

Consider two inventory policies, A and B, where

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

and

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

Then

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

but

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

So A has a lower expected value than B, but A will win almost all of the time.

Assumptions and Notation

- $\boldsymbol{X}_j = (X_{1j}, \ldots, X_{tj})$ $(j \geq 1)$ are independent observations taken from a multinomial distribution having $t \geq 2$ categories with associated unknown probabilities $\boldsymbol{p} = (p_1, \ldots, p_t)$.

- $X_{ij} = 1$ [0] if category i does [does not] occur on the jth observation.

- The (unknown) ordered p_i's are $p_{[1]} \leq \cdots \leq p_{[t]}$.

- The category associated with $p_{[t]}$ 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 $y_{[1]m} \leq \cdots \leq y_{[t]m}$.

Indifference-Zone Procedures

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

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

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

$$P\{\text{CS} \,|\, \boldsymbol{p}\} \;\geq\; P^\star \text{ whenever } p_{[t]}/p_{[t-1]} \;\geq\; \theta^\star. \qquad (5)$$

The probability in (5) depends on the entire vector $\boldsymbol{p}$ and on the number n of independent multinomial observations to be taken. The constant $\theta^\star$ is the "smallest $p_{[t]}/p_{[t-1]}$ ratio worth detecting."

Now we will consider a number of procedures to guarantee probability requirement (5).

Single-Stage Procedure $\mathcal{M}_{BEM}$ (Bechhofer, Elmaghraby, and Morse 1959):

For the given t, $\theta^\star$ and $P^\star$, find n from the table.

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

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

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

$$p_{[1]} = p_{[t-1]} = 1/(\theta^\star + t - 1) \quad \text{and} \quad p_{[t]} = \theta^\star/(\theta^\star + t - 1) \tag{6}$$

as first determined by Kesten and Morse (1959).

Example: A soft drink producer wants to find the most popular of $t = 3$ 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 Table **??** with $t = 3$, $P^\star = 0.95$, and $\theta^\star = 1.4$, we find that $n = 186$ individuals must be interviewed.

$P^\star$	$\theta^\star$	$t=2$		$t=3$		$t=4$		$t=5$	
		n	n_0	n	n_0	n	n_0	n	n_0
0.75	3.0	1	1	5	5	8	9	11	12
	2.0	5	5	12	13	20	24	29	34
	1.8	5	7	17	18	29	35	41	50
	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
0.90	3.0	7	∞	11	12	16	19	21	24
	2.0	15	15	29	34	43	53	58	71
	1.8	19	27	40	50	61	75	83	104
	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
0.95	3.0	9	11	17	20	23	26	29	34
	2.0	23	27	42	52	61	74	81	98
	1.8	33	35	59	71	87	106	115	142
	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 Size n for Procedure $\mathcal{M}_{BEM}$, and Truncation Numbers n_0 for Procedure $\mathcal{M}_{BG}$ to Guarantee (5)

A Curtailed Procedure (Bechhofer and Kulkarni 1984)

Suppose we specify the maximum total number of observations to be n. Procedure $\mathcal{M}_{BK}$ employs *curtailment* and achieves the same $P\{\text{CS}\}$ as does $\mathcal{M}_{BEM}$ with the same n. In fact,. . .

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

and

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

uniformly in $\boldsymbol{p}$, where N is the (random) number of observations to the termination of sampling.

Remark: Unlike procedure $\mathcal{M}_{BEM}$, the distance measure $\theta^\star$ does *not* play a role. The choice of the maximum number n of multinomial observations permitted can be made using criteria such as cost or availability of observations. Of course, one could also choose n to guarantee the probability requirement (5).

Procedure $\mathcal{M}_{BK}$

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

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

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

$$y_{im} \geq y_{jm} + n - m \text{ for all } j \neq i \ (1 \leq i, j \leq t). \tag{7}$$

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 $p_{[t]}$, randomizing to break ties.

Remark: The LHS of (7) 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*.

Example: For $t = 3$ 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 $t = 3$ 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 $t = 3$ 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.

Sequential Procedure with Curtailment (Bechhofer and Goldsman 1986)

Procedure $\mathcal{M}_{BG}$

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

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

At stage m, calculate the ordered category totals $y_{[1]m} \leq \cdots \leq y_{[t]m}$ and

$$z_m = \sum_{i=1}^{t-1} (1/\theta^\star)^{(y_{[t]m} - y_{[i]m})}.$$

Stop sampling at the first stage when *either*

$$z_m \leq (1 - P^\star)/P^\star \tag{8}$$

or $m = n_0$ or $y_{[t]m} - y_{[t-1]m} \geq n_0 - m$.

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

Remark: Procedure $\mathcal{M}_{BG}$ satisfies the probability requirement (5), and has the same LF-configuration as $\mathcal{M}_{BEM}$; this determines the truncation numbers given in the table.

Example: Suppose $t = 3$, $P^\star = 0.75$ 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 (8) because $z_2 = (1/3)^2 + (1/3)^2 = 2/9 \le (1 - P^\star)/P^\star = 1/3$, and we select category 2.

Example: Again suppose $t = 3$, $P^\star = 0.75$ 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 (8) because $m = n_0 = 5$ observations, and we select category 1.

Example: Yet again suppose $t = 3$, $P^\star = 0.75$ 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 (8) 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.

Example: Still yet again suppose $t = 3$, $P^\star = 0.75$ 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 third criterion in (8) tells us to stop; we select category 2.

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

Example: Suppose $t = 4$, $\theta^\star = 1.6$, $P^\star = 0.75$. The single-stage procedure $\mathcal{M}_{BEM}$ requires 46 observations to guarantee (5), whereas procedure $\mathcal{M}_{BG}$ (with a truncation number of $n_0 = 57$) has $E\{N|\text{LF}\} = 31.1$ and $E\{N|\text{EP}\} = 37.65$ for $\boldsymbol{p}$ in the LF-configuration (6) and equal-probability (EP) configuration, $p_{[1]} = p_{[t]}$, respectively.

Applications

1. Selection of the Treatment Having the Largest Location Parameter (see Bechhofer and Sobel 1958).

Suppose $W_{i1}, W_{i2}, \ldots$ is the ith of t mutually independent random samples. Assume the W_{ij} $(j \geq 1)$ are *continuous* random variables with p.d.f. $f(w - \mu_i)$ where the μ_i $(1 \leq i \leq t)$ are unknown location parameters.

Goal: Select the p.d.f. $f(w - \mu_i)$ that has the highest probability of producing the largest observation from the vector $(W_{1j}, \ldots, W_{tj})$.

Define $X_{ij} = 1$ if $W_{ij} > \max_{i' \neq i} W_{i'j}$ on trial j and $X_{ij} = 0$ if not ($1 \leq i \leq t$, $j \geq 1$). Then $(X_{1j}, \ldots, X_{tj})$ ($j \geq 1$) has a multinomial distribution with probability vector $\boldsymbol{p}$, where

$$p_i = P\{W_{i1} > W_{i'1} \ (i' \neq i;\ 1 \leq i' \leq t)\} \ (1 \leq i \leq t).$$

Let $p_{[1]} \leq \cdots \leq p_{[t]}$ denote the ordered p_i-values.

The goal of selecting the category associated with $p_{[t]}$ can be investigated using the multinomial selection procedures described in this chapter.

Remark: The p.d.f. having the highest probability, $p_{[t]}$, of producing the largest observation is the p.d.f. with the largest μ_i-value.

Warning: However, the multinomial probability requirement (5) involving the p_i's does not translate into any easily interpretable probability requirement involving the μ_i's.

Example: Use $\mathcal{M}_{BG}$ to solve the following location problem. Suppose independent observations W_{ij} are taken from $t = 3$ treatments. The p.d.f. of W_{ij} is assumed to be of the *same unknown form* $f(w - \mu_i)$ $(1 \le i \le 3)$, differing only with respect to the unknown location parameters μ_i. We want to select the treatment associated with $\mu_{[3]} = \max\{\mu_1, \mu_2, \mu_3\}$. This is equivalent to selecting the multinomial cell associated with $p_{[3]}$.

Suppose we specify $P^\star = 0.75$ and $\theta^\star = 3.0$. The truncation number $n_0 = 5$ from the table for procedure $\mathcal{M}_{BG}$ applied to the following data tells us to stop sampling if

m	w_{1m}	w_{2m}	w_{3m}	x_{1m}	x_{2m}	x_{3m}	y_{1m}	y_{2m}	y_{3m}
1	15	17	9	0	1	0	0	1	0
2	21	7	6	1	0	0	1	1	0
3	7	11	8	0	1	0	1	2	0
4	16	6	2	1	0	0	2	2	0
5	14	13	9	1	0	0	3	2	0

and select treatment 1. In doing so we are guaranteed that $P\{\text{CS}\} \geq 0.75$ whenever $p_{[3]} \geq 3p_{[2]}$.

2. A General Nonparametric Application.

Suppose we take i.i.d. vector-observations $\boldsymbol{W}_j = (W_{1j}, \ldots, W$ $(j \geq 1)$, where the W_{ij} can be either discrete or continuous random variables. For a particular vector-observation $\boldsymbol{W}_j$, suppose the experimenter can determine which of the t observations W_{ij} $(1 \leq i \leq t)$ 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 t agricultural plots using competing fertilizers.

- The smallest sample average customer waiting time based on a simulation run of each of t competing queueing strategies.

- The smallest estimated variance of customer waiting times (from the above simulations).

For a particular vector-observation $\boldsymbol{W}_j$, suppose $X_{ij} =$ 1 or 0 according as W_{ij} $(1 \leq i \leq t)$ is the "most desirable" of the components of $\boldsymbol{W}_j$ or not. Then $(X_{1j}, \ldots, X_{tj})$ $(j \geq 1)$ 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 chapter.

Example: Suppose we want to find which of $t = 3$ *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 $\min_{1 \leq i \leq 3} W_{ij}$.

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 $p_{[3]}$. This is equivalent to selecting the multinomial category associated with $p_{[3]} = \max\{p_1, p_2, p_3\}$. Specify $P^\star = 0.75$ and $\theta^\star = 3.0$. The truncation number from the table for $\mathcal{M}_{BG}$ 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.