Computer Simulation

Module 10: Comparing Systems

Dave Goldsman, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

Single-Stage Normal Means Procedure

Lesson Overview

Last Lesson: Gave a bit of background on the normal means problem.

This Time: We'll look at a specific, baby single-stage procedure.

- Just to give you a taste
- Assumes known variances
- Better procedures out there.

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 (usually from a table).

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

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

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)

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 $\mathcal{N}_\text{B}$

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

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

where $\lceil \cdot \rceil$ rounds up, and the constant $Z^{(1-P^\star)}_{k-1,1/2}$ 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.

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

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

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. $\Box$

Summary

This Time: We learned about Bechhofer's single-stage normal means procedure – and we survived the derivation of the sample size *n*!

But the procedure has limitations.

Next Time: We'll look at a bunch of more-useful normal means procedures.

Computer Simulation

Module 10: Comparing Systems

Dave Goldsman, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

Normal Means Extensions

GTx

Lesson Overview

Last Lesson: Bechhofer’s famous single-stage normal means procedure.

Almost 1100 Google Scholar citations!

This Time: But the single-stage procedure has limitations… so we’ll take a high-level look at more-useful procedures.

No math necessary in this lesson!

Good News ☺

Bechhofer's single-stage procedure is sooo easy and intuitive.

- Take a bunch of observations from each population.
- Select as best the one with the largest sample mean.
- Required sample size just requires a table look-up.
- Procedure is actually kind of robust against certain assumption violations.

But... ☹

Bechhofer's procedure has some major drawbacks.

- It assumes that the variances are known and common – cRazy talk!
- It's conservative since sample size is for the least-favorable configuration of the means.
- Requires the competing populations to be independent.
- Have to be careful when using it for simulation data.

What to Do?

Any real-life problem will involve unknown and unequal variances.

- Rinott's two-stage procedure
  - Estimates variances in stage 1
  - Determines how many observations to take in stage 2
- Kim and Nelson sequential procedure
  - Estimates variances in stage 1
  - Samples in multiple stages
  - Eliminates poor populations as sampling proceeds.

What to Do?

These procedures obviously require i.i.d. normal observations within each system.

- Of course, stuff like consecutive waiting times is dependent and nonnormal.
- No need to panic!
  - Batch means are approximately normal for large batch size.
  - Batch means are approximately independent.

What to Do?

Ordinarily, the competing systems should be independent.

- But various procedures out there if you know that the systems are correlated.
- Useful in simulations where you can induce positive correlation between systems.
- If handled correctly, correlation is helpful in distinguishing the best.

Summary

This Time: The single-stage procedure was too simplistic, so we looked at ways to apply normal means procedures in practical situations – not so bad!

But what if the competing systems aren't normal?

Next Time: The Bernoulli selection problem. Which Bern has the largest *p*?

Computer Simulation

Module 10: Comparing Systems

Dave Goldsman, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

Bernoulli Probability Selection

Lesson Overview

Last Lesson: Wrapped up our discussion on the normal means selection problem.

This Time: Selecting the Bernoulli population with the largest success parameter.

Lots of applications in simulation, medical trials, etc.

Bernoulli Selection – Introduction

Select the Bernoulli population with the largest success parameter.

Examples:

- Which anti-cancer drug is most effective?
- Which simulated system is most likely to meet design specs?
- Which (*s*,*S*) inventory policy has the highest profit probability?

There are 100's of such procedures. Highlights:

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

Again use the *indifference-zone* approach.

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

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

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

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 $\mathcal{B}_{\text{SH}}$ (Sobel and Huyett 1957)

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

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

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.

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 $\mathcal{B}_{\text{SH}}$ 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. $\square$

Summary

This Time: We introduced the Bernoulli selection problem; set up the necessary notation; and gave a single-stage procedure.

Next Time: A couple of better Bernoulli procedures involving an easy “curtailment” trick and a simple sequential procedure.

Computer Simulation

Module 10: Comparing Systems

Dave Goldsman, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

Bernoulli Extensions

Lesson Overview

Last Lesson: Introduced the Bernoulli selection problem and discussed a single-stage procedure.

This Time: Better procedures that save observations.

We’re off to the Bern Unit!

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 $\mathcal{B}_{\text{BKS}}$ (BKS 1968)

New Prob Requirement: For specified $(P^\star, \theta^\star)$ with $1/k < P^\star < 1$ and $\theta^\star > 1$, we require $P(\text{CS}) \geq P^\star$ 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!

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 $\mathcal{B}_{\text{BKS}}$.

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. $\square$

Summary

This Time: Introduced some efficient tricks for the Bernoulli selection problem.

Next Lesson: We'll introduce the multinomial selection problem – choose the most-probable category.

Computer Simulation

Module 10: Comparing Systems

Dave Goldsman, Ph.D.

Professor

Stewart School of Industrial and Systems Engineering

Multinomial Cell Selection

GTx

Lesson Overview

Last Lesson: Finished off the Bernoulli selection problem with a couple of nice efficiency tricks.

This Time: The multinomial selection problem – find the most-probable category.

Lots of applications – surveys, simulations, etc.

Multinomial Selection – Intro

Select the multinomial category having the largest probability.

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.

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.

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

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

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

Outcome (red, blue, violet)	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

$P(\text{CS}) = 0.6042$

Can increase $P(\text{CS})$ by increasing n.

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

Consider two inventory policies, A and B, where

$$
\begin{aligned}
\text{Profit from } A &= \$5 \text{ with probability } 1 \\
\text{Profit from } B &= \begin{cases} \$0 & \text{with probability } 0.99 \\ \$1000 & \text{with probability } 0.01. \end{cases}
\end{aligned}
$$

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

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

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

Summary

This Time: Introduced the multinomial selection problem with some motivational examples – but no procedures.

Next Lesson: Some notation + an actual procedure to select the most-probable cell!

Computer Simulation

Module 10: Comparing Systems

Dave Goldsman, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

Multinomial Procedure + Extensions

Lesson Overview

Last Lesson: Introduced the multinomial selection problem along with some motivational examples.

This Time: Single-stage procedure for selecting the most-probable cell + a couple of extensions.

Time is winding down!

Assumptions and Notation for Selection Problem

- $\boldsymbol{X}_j = (X_{1j}, \ldots, X_{kj})$ $(j \geq 1)$ 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 $Y_{[1]m} \leq \cdots \leq Y_{[k]m}$.

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

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

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

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

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

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 $\mathcal{M}_{\text{BEM}}$ achieves $P(\text{CS}) \geq P^\star$ when the cell probabilities $\boldsymbol{p}$ are in the *least-favorable* (LF) configuration (Kesten and Morse 1959),

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

Tick Tock…

$P^\star$	$\theta^\star$	$k=2$	$k=3$	$k=4$	$k=5$
	2.0	5	12	20	29
	1.8	5	17	29	41
0.75	1.6	9	26	46	68
	1.4	17	52	92	137
	1.2	55	181	326	486
	2.0	15	29	43	58
	1.8	19	40	61	83
0.90	1.6	31	64	98	134
	1.4	59	126	196	271
	1.2	199	437	692	964
	2.0	23	42	61	81
	1.8	33	59	87	115
0.95	1.6	49	94	139	185
	1.4	97	186	278	374
	1.2	327	645	979	1331

Sample Size n for $\mathcal{M}_{\mathrm{BEM}}$ to Guarantee (3)

Tick Tock…

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. $\square$

Tick Tock…

Extensions

Curtailed procedure: Stop sampling when the guy in second place can’t win.

Sequential procedures: Sample one at a time and stop when the winner is clear.

Real-world examples: Easy to apply this stuff – each observation is a simulation run.

Tick Tock!!

Summary

This Time: Looked at the single-stage most-probable multinomial cell selection procedure + a couple of extensions.

This completes the module.

This completes the course.

The Bell Tolls for Thee!

Goodbye Dear Students!