Selective Assembly in Manufacturing: Statistical Issues and Optimal Binning Strategies

David Mease
Department of Statistics
Wharton School, University of Pennsylvania
Philadelphia, PA 19104-6340
dmease@wharton.upenn.edu

Vijayan N. Nair
Department of Statistics
University of Michigan
Ann Arbor, MI 48109-1092
vnn@umich.edu

Agus Sudjianto
Risk Management
Bank of America
Charlotte, NC 28255-0001
Agus.Sudjianto@bankofamerica.com

Abstract

Selective assembly is a cost-effective approach for reducing the overall variation and thus improving the quality of an assembled product. In this process, components of a mating pair are measured and grouped into several classes (bins) as they are manufactured. The final product is assembled by selecting the components of each pair from appropriate bins to meet the required specifications as closely as possible. This approach is often less costly than tolerance design using tighter specifications on individual components. It leads to high-quality assembly using relatively inexpensive components. In this paper, we describe the statistical formulation of the problem and develop optimal binning strategies under several loss functions and distributional assumptions. Optimal schemes under absolute and squared error loss are studied in detail. The results are compared with two commonly used heuristic schemes. We consider situations where only one component of the mating pair is binned as well as cases where both components are binned.

KEY WORDS: Match Gaging; Optimal Partitioning; Tolerance Design; Variation Reduction.

1 Introduction

The quality and performance of an assembled product depends critically on the dimensional variation of its component parts. In fact, part dimensional variation is one of the major sources of quality problems in the automobile industry (Ceglarek and Shi, 1995). In most applications, the overall tolerance of the assembly is determined by the sum of the individual component tolerances. Thus, the components need to be manufactured at a much higher level of precision in order to meet the overall assembly tolerance. This can be very costly and sometimes even infeasible. Thesen and Jantayavichit (1999) describe an application involving scroll compressors where the individual shells that form the compression cavity cannot be manufactured with sufficient precision using standard high-volume machining processes.

Selective assembly (sometimes also called match gaging) is a cost-effective alternative in these situations. It can be used to achieve high-precision assembly from relatively low-precision components. Selective assembly focuses on the fit between mating parts rather than the absolute dimension of each component. In this approach, the mating components are measured, sorted by dimension and binned into groups prior to the assembly process. The final product is assembled by selecting the components from appropriate bins to achieve optimal assembly dimensions (e.g. clearance). We illustrate the procedure through two examples.

Figure 1 is an example from the automobile industry and pictures a conventional internal combustion engine valvetrain assembly involving a camshaft, a valve and a tappet. In the figure, the camshaft is shown in the “nose up” position, but as the camshaft rotates the nose of the camshaft repeatedly makes contact with the tappet and pushes it down for a period of time causing the valve to open. The desired value for the clearance between the bottom of the camshaft and the top of the tappet is 300μm. However, due to variation in both the distance Y from the bottom of the camshaft to the top of the valve as well as variation in the width X of the tappet, the actual clearance $Y - X$ will vary from this value. Deviation from the target clearance means that, for each cycle, the valve will remain open for a longer or shorter period of time than specified by the design, causing engine performance to suffer. We can use selective assembly to reduce this variation. In this process, tappets are separated into a number of bins based on their widths. As each camshaft and valve assembly comes through the production line, a robot measures the distance Y and then selects a tappet with width X from the correct bin to complete the assembly with the goal of creating the target clearance of 300μm. In this example, only one component of the mating pair (the tappet) is binned.

The second example involves a measurement arm attached to a shaft (Fang and Zhang

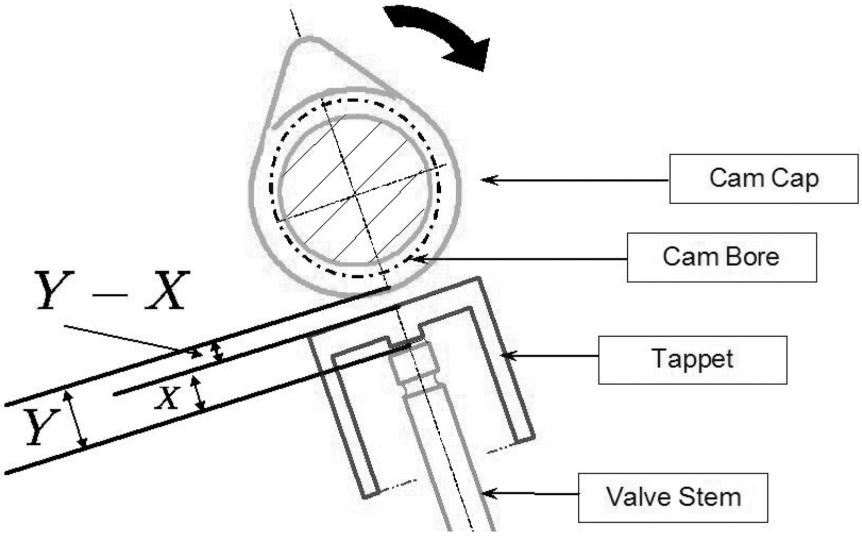


Figure 1: Camshaft, Tappet and Valve Assembly

(1995)). Both the diameter of the shaft (X) and the diameter of the hole in the measurement arm into which the shaft is inserted (Y) possess some degree of variation. This leads to variation in the clearance ($Y - X$), which is a critical characteristic since the measurement arm is designed to slide along the shaft. If the clearance is smaller than 12μm, it becomes difficult to manually slide the arm along the shaft; on the other hand, if the clearance is greater than 20μm the concentricity accuracy of the measurement arm becomes unacceptable. In selective assembly, the measurement arms and the shafts are grouped into a number of different bins after being manufactured. The final assemblies are made by matching arms and shafts from corresponding bins. In this way, selective assembly matches larger (smaller) shafts with arms that have larger (smaller) holes, resulting in less variability in the clearance than if random assembly were used.

In both these examples, the assembly dimension of interest was the clearance $Y - X$ between two mating parts. However, selective assembly can also be used in applications where the goal is to reduce the variation in the total height $Y + X$ of an assembly consisting of two components stacked on top of one another. Such situations can be translated into problems dealing with the clearance by replacing X by $-X$. So, without loss of generality, we will restrict the focus of the paper to problems dealing with clearance $Y - X$.

Selective assembly is used extensively within the automobile industry involving traditional applications such as pistons and cylinders as well as crankshafts and bearings. As noted in Pugh (1992), it may be even more widely used in the future due to recent trends in manufacturing industries. The increasing use of flexible manufacturing systems, where the same process is used to manufacture many different parts, is likely to result in larger component variation. At the same time, advances in measurement technology and ease of automated inspection make it easier to inspect, measure, and bin the components prior to assembly.

The design of efficient selective assembly processes involve a number of interesting statistical and optimization issues. In this paper, we focus on binning strategies and optimal schemes. Other issues such as optimizing inventory buffers will be described briefly in the final section. Most of the binning strategies proposed in the literature for selective assembly involve one of two heuristic methods: equal width or equal area partitioning schemes. Equal width schemes, as the name suggests, partition the dimensional distributions of the components so that the bins have equal widths (see Pugh (1986) and Thesen and Jantayavichit (1999)). In equal area schemes (Pugh, 1992), the distributions are divided into classes so that the bins have equal area or equal probabilities. There has been very little work on the development of optimal schemes. Kwon, Kim, and Chandra (1999) is the only paper that we know of that has studied optimal partitioning schemes. They restrict attention to identical normal distributions for both components in the pair and develop schemes that minimize squared error loss. There has not been any work on other types of loss functions, other distributions, or situations where the component distributions are different.

In this paper, we study optimal binning strategies for different loss functions and distributions and compare them to the heuristic methods based on equal area and equal width. In particular, we develop optimal strategies under squared and absolute error loss functions and discuss algorithms for finding the optimal partitions for general distributions. We also provide results on the existence and uniqueness of the optimal partitions in the squared error loss case.

It turns out that the solution in the squared error loss case is *mathematically* equivalent to that for an optimal partitioning problem that has been studied in the literature. This is a "one-sided" partitioning problem where the goal is to find the best $n-$point discrete distribution to approximate a given distribution. The problem can also be cast as finding the best piecewise constant approximation to a given function. Eubank (1988) provides an excellent review of the problem and applications. See also Tarpey and Flury (1996) for related discussion on self-consistent and principal points. The connection between our problem and this one-sided partitioning problem in the squared error loss case is discussed in Section 3.4. Interestingly, there is no relationship between the two problems for other loss functions.

2 Models, Notation and Assumptions

We begin with the case in which both components of the pair are binned, as in the example involving the measurement arm and shaft. Let X and Y be continuous random variables denoting the relevant dimensions of the two respective components so that the clearance between the components is given by $Y - X$. Let n denote the number of bins and let $(x_0, x_1, ..., x_n)$ and $(y_0, y_1, ..., y_n)$ be the partition limits for X and Y respectively. We will fix x_0 and y_0 to be the lower end points for the support of the X and Y distributions respectively and x_n and y_n to be the upper end points. In practice, these will be finite, but we will entertain distributions with infinite support for the sake of generality. In this case, x_0 or y_0 can be $-\infty$ and x_n or y_n can be $+\infty$. Components having X values between x_{i-1} and x_i are matched with components having Y values that are between y_{i-1} and y_i. We denote $\mathrm{P}(x_{i-1} < X \leq x_i)$ and $\mathrm{P}(y_{i-1} < Y \leq y_i)$ by p_i^X and p_i^Y respectively. Further, let X_i and Y_i be the random variables X and Y conditioned to be on the intervals $(x_{i-1}, x_i]$ and $(y_{i-1}, y_i]$ respectively. We will assume initially that $p_i^X = p_i^Y$ for all i. This assumption ensures that there is no surplus over time of either the X or Y component in any one bin. We will relax this assumption in Section 5.

Let $L(Y - X - \tau)$ be the loss incurred when the target clearance is τ and we use components with dimensions X and Y. Then, the expected loss using the above binning scheme is

$$\sum_{i=1}^{n} p_i \mathrm{E} L(Y_i - X_i - \tau). \tag{1}$$

The goal is to find the set of partition limits $(x_1, x_2, ..., x_{n-1})$ and $(y_1, y_2, ..., y_{n-1})$ that minimize (1) for a fixed n. Throughout, we assume that the processes are on target on the average, i.e., $\mathrm{E}(Y - X) = \tau$, so without loss of generality we can let $\tau = 0$. Further, by appropriately re-centering the distributions of X and Y, we can also assume that $E(X) = E(Y) = 0$.

We will study squared and absolute error loss functions in detail. The squared error loss is

$$\sum_{i=1}^{n} \mathrm{E}(Y_i - X_i)^2 p_i \tag{2}$$

and the absolute error loss is

$$\sum_{i=1}^{n} \mathrm{E}|Y_i - X_i| p_i. \tag{3}$$

For cases in which only one of the two components is binned such as the first example involving the tappets, we will let X be the relevant dimension of the component that is binned and Y be the dimension of the component that is not binned. The X components from the ith bin (X is between x_{i-1} and x_i) will be used when Y is between, say, y_{i-1} and y_i. Thus, the binning of the X component *induces* a partitioning of the Y distribution, and as such this problem can be treated as if both of the two components are binned. The one exception to this concerns the restriction that p_i^X is equal to p_i^Y for all i. For applications in which the binned component is cheap relative to the other component (as might be the case with the tappets) the cost of wasted parts due to a surplus may be small relative to the cost associated with the overall quality of the assembly. In these cases, the restriction that p_i^X is equal to p_i^Y for all i is not necessary. We will examine the consequences of relaxing this restriction in Section 5.

3 Optimal Binning Strategies Under Squared Error Loss

The development of optimal strategies for squared error loss is complicated in general. We will first consider two special cases in which we can characterize the nature of the solutions. Section 3.1 considers the case in which X and Y have the same distribution and Section 3.2 extends these results to the case when the distributions differ only by a scale parameter. Section 3.3 discusses the general problem in which the X and Y distributions are arbitrary.

3.1 Identical Distributions for Component Pairs

In some cases, it is reasonable to assume that the distributions for X and Y components are the same (at least up to differences in the mean). This, together with the assumption that p_i^X is equal to p_i^Y for all i stated earlier, implies that $(x_1, x_2, ..., x_{n-1}) = (y_1, y_2, ..., y_{n-1})$. Thus, the problem of minimizing (2) simplifies substantially as only one set of partition limits $(x_1, x_2, ..., x_{n-1})$ must be found. (Recall that the end points x_0, y_0, x_n and y_n are fixed.)

To find the set of partitions to minimize (2), note first that

$$\sum_{i=1}^{n} \mathrm{E}(Y_i - X_i)^2 p_i = \mathrm{E}(Y^2) + \mathrm{E}(X^2) - 2\sum_{i=1}^{n} \mathrm{E}(Y_i X_i) p_i. \tag{4}$$

The first two terms do not depend on the partitions, so it is sufficient to maximize the last term

$$\sum_{i=1}^{n} \mathrm{E}(Y_i X_i) p_i. \tag{5}$$

Since X_i and Y_i are iid, this is simply

$$\sum_{i=1}^{n} [\mathrm{E}(X_i)]^2 p_i. \tag{6}$$

Letting f denote the density of X (and Y) and F the corresponding cumulative distribution function, we can rewrite (6) as

$$\sum_{i=1}^{n} \frac{\left(\int_{x_{i-1}}^{x_i} x f(x) dx\right)^2}{F(x_i) - F(x_{i-1})}. \tag{7}$$

Taking the derivative with respect to each x_i $(1 \leq i \leq n-1)$ gives

$$2f(x_i)[\mathrm{E}(X_i) - \mathrm{E}(X_{i+1})][x_i - \frac{\mathrm{E}(X_i) + \mathrm{E}(X_{i+1})}{2}].$$

Since the optimal partition limits correspond to sign changes in this derivative, they must satisfy the equations

$$x_i = \frac{\mathrm{E}(X_i) + \mathrm{E}(X_{i+1})}{2}, 1 \leq i \leq n-1. \tag{8}$$

Because the conditional means $\mathrm{E}(X_i)$ depend on the partition limits x_i, (8) does not yield a closed form expression for the optimal partitions. However, any of the standard algorithms can be used to find these stationary values and the global optimum. The following is a simple fixed-point functional iteration algorithm:

(1) Begin with an initial set of partition limits $(x_1^0, x_2^0, ..., x_{n-1}^0)$.

(2) Using these initial partitions, compute the conditional means $\mathrm{E}(X_1)^0, \mathrm{E}(X_2)^0, ..., \mathrm{E}(X_n)^0$.

(3) Compute

$$x_i^1 = \frac{\mathrm{E}(X_i)^0 + \mathrm{E}(X_{i+1})}{2}^0, 1 \leq i \leq n-1.$$

(4) Go back to Step 2 using $(x_1^1, x_2^1, ..., x_{n-1}^1)$ in place of $(x_1^0, x_2^0, ..., x_{n-1}^0)$ and iterate until convergence.

Expression (8) as well as the above algorithm were obtained by Kwon, Kim and Chandra (1999) for the case with normal distributions. However, as the above development shows, the results hold for any general distribution F.

It can be verified that for any $n > 1$ there exists a set of partition limits which minimizes (2). The essential idea of a proof is as follows. The result is clear for distributions with finite (compact) support since the criterion is a continuous function of the partition limits. Further, the optimum limits must fall in the interior of the support; otherwise there exists a set of partitions with $(n-1)$ points that is better than those with n points which is a contradiction. The proof can then be extended to distributions with infinite support by an approximation argument. We omit the details.

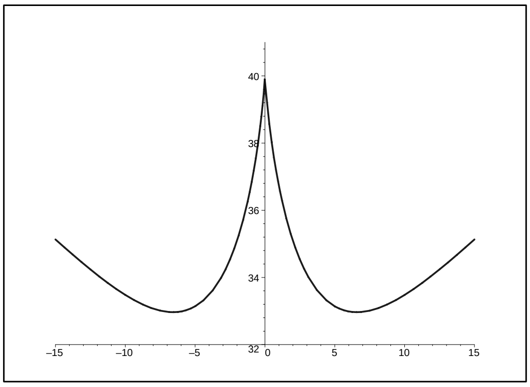
The solution to (8) is not necessarily unique, so there is no guarantee that a solution is the global optimum. (See also Tarpey (1994), Eubank (1988) and Trushkin (1982)). In fact, there are multiple solutions in general as shown by the following two examples which also illustrate several other interesting features.

Example 1: Consider a symmetric continuous distribution given by the *double Weibull* distribution with density

$$f(x) = \frac{1}{2}\frac{\beta}{\eta}\left(\frac{|x|}{\eta}\right)^{\beta-1} e^{-(\frac{|x|}{\eta})^{\beta}}.$$

Take $\eta = 1$ and $\beta = 1/2$ in the above. For $n = 2$, let $(-\infty, x_1]$ and (x_1, ∞) be the partitions. The left panel in Figure 2 displays the expected squared error loss as a function of x_1. This plot shows that that the (obvious) symmetric solution $x_1 = 0$ is in fact a local maximum, while the global minimum is achieved by either of the asymmetric solutions: $x_1 = \pm 6.584$ (to three decimals).

The right panel shows the identity function (left-hand side of (8)) and the curve $\frac{E(X_1)+E(X_2)}{2}$ (right-hand side of (8)). The crossings of the two functions are the stationary points of (8). Note that the two global minima correspond to a "downcrossing" (i.e., the curve crosses the identity function from above) while the local maximum is an "upcrossing." In this example, the fixed-point functional iteration algorithm discussed earlier will never converge to the symmetric solution ($x_1 = 0$) unless the initial value is zero. It appears that this is true in general, i.e., upcrossings of the two curves (right panel in Figure 2) are local maxima and the fixed-point algorithm will not converge to these points.



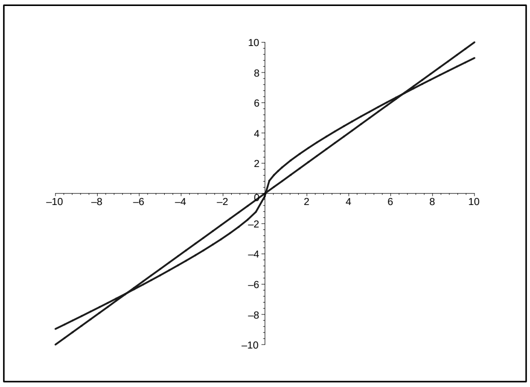


Figure 2: Squared Error Loss as a Function of x_1 for Example 1 (left panel) and Crossing of the Identity Function and $\frac{E(X_1)+E(X_2)}{2}$ for Example 1 (right panel)

Example 2: In the first example, the symmetric solution of $x_1 = 0$ corresponded to a local maximum. This example, which deals with three partitions, shows that there can be several symmetric solutions and these can be local minima or maxima. Consider the symmetric distribution with

$$f(x) = \begin{cases} 0 & |x| > 4 \\ \frac{1}{60} & 1 \le |x| < 4 \\ \frac{9}{20} & |x| < 1. \end{cases}$$

For $n = 3$, it can be verified directly that there are multiple symmetric solutions to equation (8), given by $(x_1, x_2) = (-\frac{20-\sqrt{22}}{27}, \frac{20-\sqrt{22}}{27})$, $(-\frac{20+\sqrt{22}}{27}, \frac{20+\sqrt{22}}{27})$ and $(-\frac{4}{3}, \frac{4}{3})$. The global minimum is achieved by the asymmetric solution $(x_1, x_2) = (-1.605, 0.219)$ (to three decimals). The curve in Figure 3 is the expected squared error loss for all symmetric partitions as a function of x_2. Note that the solutions $(-\frac{20-\sqrt{22}}{27}, \frac{20-\sqrt{22}}{27})$ and $(-\frac{4}{3}, \frac{4}{3})$ are local minima while $(-\frac{20+\sqrt{22}}{27}, \frac{20+\sqrt{22}}{27})$ is a local maximum. The global minimum corresponds to the horizontal line, which is achieved by the asymmetric solution. If we start with initial values that are close to -4 and 4, then the fixed point iteration algorithm will stop at $(-\frac{4}{3}, \frac{4}{3})$ and we will not find the global minimum.

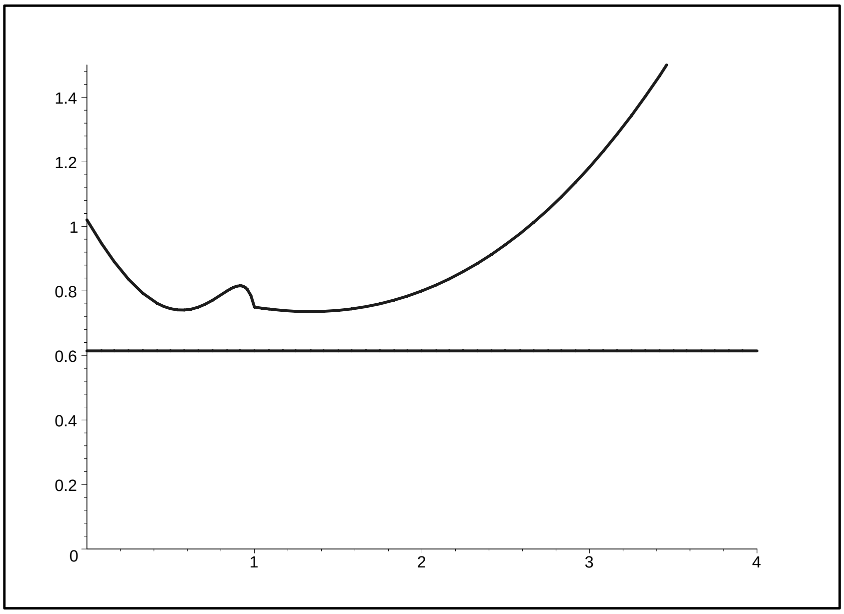


Figure 3: Squared Error Loss for Symmetric Partitions as a Function of x_2 (Example 2)

The above examples demonstrate that a solution to (8) may correspond to a local minimum or maximum and the global minimum can be overlooked. Proposition 1 establishes conditions under which the solution to (8) is unique and is the global minimum.

Proposition 1: Suppose the density f is positive on some interval (A, B) where $-\infty \le A < B \le \infty$ and zero elsewhere and

$$\frac{\int_t^{t+h} x f(x)dx}{\int_t^{t+h} f(x)dx} - t \tag{9}$$

is non-increasing in $t \in (A, B)$ for all h in $(0, \infty)$. Then, there exists at most one solution to (8).

After obtaining this result, we discovered that Trushkin (1982) had also established the result under similar conditions for the squared error loss case. The statement of his conditions appears more complicated, but it can be simplified to the conditions of Proposition 1. For the sake of completeness, our proof is included in the Appendix.

The condition in (9) can be interpreted as follows. We can think of $\frac{\int_t^{t+h} xf(x)dx}{\int_t^{t+h} f(x)dx}$ as the conditional mean of X in the interval $(t, t+h)$. For $h = \infty$, expression (9) is the mean residual life that arises in reliability. For finite h, this can be viewed as the mean residual life if X is restricted to the interval $(t, t+h)$. Proposition 1 requires that this (conditional) mean residual life is non-increasing for every h. Thus, it is stronger than the non-increasing mean residual life condition. The following lemma shows that the conditions of Proposition 1 are satisfied if the density f is strongly unimodal. The proof is deferred to the Appendix.

Lemma 1: If the density $f(x)$ is strongly unimodal (i.e., $\log f(x)$ is concave), then (9) is non-increasing in t for all h in $(0, \infty)$.

Remarks: Kwon, Kim, and Chandra (1999) discuss the existence and uniqueness of the solutions to (8) for the special case of the normal distribution, but their proof is incorrect. One implication of this uniqueness result is that, for a symmetric density that satisfies these conditions, any solution must lead to a symmetric partition. Our examples, in which asymmetric partitions were optimal for symmetric distributions, do not satisfy the conditions.

We computed the optimal partition limits for several different distributions using numerical methods. Tables 1 and 3 give results for the normal distribution, the double exponential distribution and the logistic distribution, all standardized to have variance one and then truncated at ± 3. Table 2 gives results for the standard normal distribution truncated at ± 2. The optimal partition limits for the uniform distribution reduce to equal width (or equal area) schemes. Figure 4 compares the relative efficiencies of the optimal binnings for these distributions to the heuristic methods of equal area and equal width partitioning.

Table 1: Standard Normal Distribution Truncated at -3 and 3 (Squared Error Loss)

n	Optimal Partition Limits (Only the nonnegative values are given since all partitions are symmetric.)							Expected Squared Difference from Target	Percentage Savings Over Equal Area Partitioning	Percentage Savings Over Equal Width Partitioning
1	-							1.947	-	-
2	0.000							.695	0%	0%
3	0.604							.358	7.91%	29.56%
4	0.000	0.964						.218	15.47%	31.01%
5	0.375	1.215						.146	21.96%	31.75%
6	0.000	0.643	1.405					.105	27.44%	31.83%
7	0.273	0.850	1.555					.079	32.07%	31.71%
8	0.000	0.486	1.017	1.677				.061	36.03%	31.53%
9	0.215	0.659	1.154	1.779				.049	39.44%	31.35%
10	0.000	0.391	0.804	1.271	1.866			.040	42.41%	31.19%
11	0.177	0.539	0.928	1.372	1.940			.033	45.02%	31.04%
12	0.000	0.327	0.667	1.035	1.460	2.005		.028	47.32%	30.91%
13	0.151	0.457	0.778	1.130	1.538	2.063		.024	49.37%	30.81%
14	0.000	0.281	0.570	0.877	1.214	1.607	2.113	.021	51.21%	30.71%
15	0.131	0.396	0.671	0.965	1.289	1.669	2.159	.018	52.87%	30.63%

Table 2: Standard Normal Distribution Truncated at -2 and 2 (Squared Error Loss)

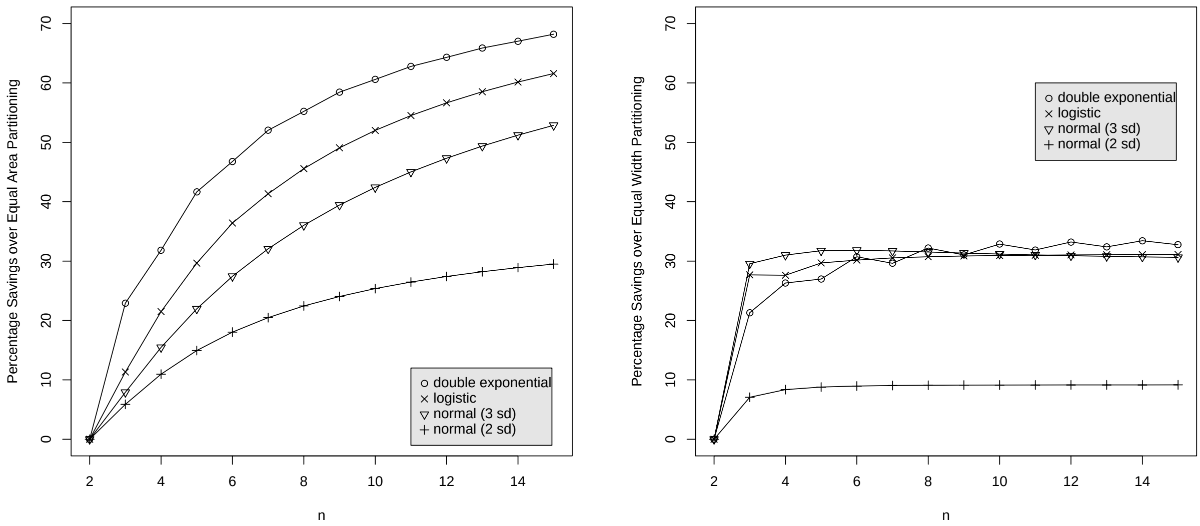
n	Optimal Partition Limits (Only the nonnegative values are given since all partitions are symmetric.)							Expected Squared Difference from Target	Percentage Savings Over Equal Area Partitioning	Percentage Savings Over Equal Width Partitioning
1	-							1.547	-	-
2	0.000							.503	0%	0%
3	0.532							.244	5.88%	7.07%
4	0.000	0.829						.142	10.96%	8.36%
5	0.323	1.022						.093	14.94%	8.79%
6	0.000	0.546	1.158					.065	18.04%	8.96%
7	0.232	0.713	1.261					.048	20.48%	9.05%
8	0.000	0.409	0.843	1.341				.037	22.45%	9.09%
9	0.181	0.550	0.947	1.404				.030	24.04%	9.12%
10	0.000	0.327	0.666	1.033	1.457			.024	25.37%	9.13%
11	0.148	0.448	0.763	1.106	1.501			.020	26.47%	9.14%
12	0.000	0.272	0.551	0.846	1.168	1.538		.017	27.41%	9.15%
13	0.125	0.378	0.640	0.918	1.221	1.570		.014	28.21%	9.15%
14	0.000	0.233	0.471	0.718	0.980	1.268	1.597	.012	28.90%	9.15%
15	0.109	0.328	0.552	0.786	1.036	1.310	1.622	.011	29.50%	9.16%

Table 3: Nonnegative Optimal Partition Limits

n	Normal	Double Exponential	Logistic
2	0.000	0.000	0.000
3	0.604	0.622	0.596
4	0.000 0.964	0.000 0.961	0.000 0.966
5	0.375 1.215	0.358 1.260	0.366 1.231
10	0.000 0.391 0.804 1.271 1.866	0.000 0.345 0.758 1.272 1.958	0.000 0.378 0.788 1.273 1.919
15	0.131 0.396 0.671 0.965 1.289 1.669 2.159	0.112 0.350 0.616 0.923 1.281 1.713 2.258	0.125 0.381 0.650 0.954 1.283 1.689 2.218

For all distributions studied, the optimal schemes give substantial improvement over both equal area and equal width binning. As seen in Figure 4, the percentage improvement of the optimal binning over equal width binning seems to increase as the number of bins is increased, while for equal width binning this remains fairly constant for $n > 4$. Equal width binning is more efficient than equal area binning. As a general rule, for all distributions given in the tables, the bins where the density is high are not as wide as those where the density is

Figure 4: Percentage Savings Versus Equal Area and Equal Width Partitioning



low. This helps to explain the superiority of the optimal binning method over equal width binning. Under optimal binning, components which occur more frequently are binned more finely than those which occur less frequently. The equal area method also attempts to do this, but it appears to overdo this and hence leads to extremely suboptimal partition limits. Table 3 suggests that the optimal partition limits for the three distributions listed are reasonably close. The right panel in Figure 4 shows that the savings in efficiency over equal width partitioning are also reasonably close for these three distributions. However, there is considerable difference in savings over equal area schemes (left panel of Figure 4).

3.2 Distributions of Component Pairs Differ by a Scale Parameter

The results in Section 3.1 can be easily extended to the case where the Y and X distributions belong to the same scale family, In this case, $\mathrm{P}(Y \leq y) = \mathrm{P}(X \leq \sigma y)$ for all y for some scale parameter $\sigma > 0$. This along with the restriction that p_i^Y is equal to p_i^X for all i implies that $(x_1, x_2, ..., x_{n-1}) = (\sigma y_1, \sigma y_2, ..., \sigma y_{n-1})$ and that $(\mathrm{E}(X_1), \mathrm{E}(X_2), ..., \mathrm{E}(X_n)) = (\sigma \mathrm{E}(Y_1), \sigma \mathrm{E}(Y_2), ..., \sigma \mathrm{E}(Y_n))$. Using this we can rewrite (2) as

$$\sum_{i=1}^{n} \mathrm{E}(Y_i - X_i)^2 p_i = \mathrm{E}(Y^2) + \sigma^2 \mathrm{E}(Y^2) - 2\sigma \sum_{i=1}^{n} [\mathrm{E}(Y_i)]^2 p_i. \tag{10}$$

The first two terms on the right hand side do not depend on the partitions, so we need to find only the partition limits $(y_1, y_2, ..., y_{n-1})$ to maximize the third expression. These can be obtained using (8) as in the last section. The partitions for the X distribution can then be obtained using the relationship $(x_1, x_2, ..., x_{n-1}) = (\sigma y_1, \sigma y_2, ..., \sigma y_{n-1})$.

Since the optimal $(y_1, y_2, ..., y_{n-1})$ given by (8) do not depend on σ, we can minimize (10) as a function of σ and $(y_1, y_2, ..., y_{n-1})$ simultaneously. This leads to the optimal value of

$$\sigma = \frac{\sum_{i=1}^{n}[E(Y_i)]^2 p_i}{\mathrm{E}(Y^2)}. \tag{11}$$

Note that this value is not necessarily equal to one, so in the optimal case the two component distributions are not the same. Pugh (1992) discusses selective assembly with components that have unequal variances. He suggests truncating the distribution with larger variance so that the variances are approximately equal. The above result suggests that this may not be the optimal strategy.

3.3 Arbitrary Distributions for Component Pairs

When the distributions of Y and X are arbitrary, it is more difficult to characterize the optimal bins along the lines of Sections 3.1 and 3.2. In general, we have to resort to the use of numerical methods to find the optimal partitions. But there are special cases where the optimal bins can still be derived analytically,

We still assume that $p_i^Y = p_i^X = p_i$ for all i. Let F_Y and F_X be the cumulative distribution functions of Y and X respectively, and let f_Y and f_X denote the corresponding densities. Note that each x_i can be expressed as $F_X^{-1}(F_Y(y_i))$ so we can rewrite (5) as

$$\sum_{i=1}^{n} \frac{\left(\int_{F_X^{-1}(F_Y(y_{i-1}))}^{F_X^{-1}(F_Y(y_i))} x f_X(x)dx\right)\left(\int_{y_{i-1}}^{y_i} y f_Y(y)dy\right)}{F_Y(y_i) - F_Y(y_{i-1})}. \tag{12}$$

In some special cases, we can differentiate the above expression and obtain formulas or iterative algorithms for the optimal partitions, analogous to those in Section 3.1. For example, if X has a uniform distribution over $(-1, 1)$ we can write (12) as

$$\sum_{i=1}^{n} \left(\int_{y_{i-1}}^{y_i} y f_Y(y)dy\right)(F_Y(y_i) + F_Y(y_{i-1}) - 1).$$

Differentiating with respect to the y_i and setting the derivative to zero yields the following expression for the stationary values:

$$y_i = \frac{\mathrm{E}(Y_i)p_i + \mathrm{E}(Y_{i+1})p_{i+1}}{p_i + p_{i+1}}, 1 \leq i \leq n-1. \tag{13}$$

We can use (13) in a recursive algorithm, analogous to the one in Section 3.1, to find the optimal partition limits $(y_1, y_2, ..., y_{n-1})$. The values for $(x_1, x_2, ..., x_{n-1})$ can then be obtained by equating the p_i^X to p_i^Y for all i.

In general, however, one has to resort to numerical methods to determine the optimal partition limits. One approach is to simply search the $(n-1)-$dimensional region of possible values for $(y_1, y_2, ..., y_{n-1})$. At each point, selected values for $(x_1, x_2, ..., x_{n-1})$ are obtained from the requirement that p_i^Y is equal to p_i^X for all i, and the value of (2) could then be evaluated numerically. If both Y and X have symmetric distributions and one is willing to restrict attention to symmetric partitions, the dimensionality of the region can be reduced to $(n-2)/2$ when n is even and $(n-1)/2$ when n is odd.

When n is large, the following heuristic approach could be used. Choose one of the two distributions (say F_Y) and apply the algorithm from Section 3.1 to find the partition limits for that distribution. The partition limits for the second distribution are then obtained from the requirement that p_i^Y is equal to p_i^X for all i. Since the resulting two sets of partition limits, $(y_1, y_2, ..., y_{n-1})$ and $(x_1, x_2, ..., x_{n-1})$, would depend on which of the two distributions is initially chosen, one can try this method for both of the two distributions and choose the partition limits that give the smaller value of (2). Alternatively, one can consider using the averages of either the Y or X partition limits resulting from applying this method for each of the two distributions.

To compare the usefulness of these various methods, we return to the case where Y has a standard normal distribution and X is distributed uniformly on $(-1, 1)$. Column A in Table 4 gives the squared error loss (value of (2)) based on the optimal partition limits found using the algorithm based on (13). The other four columns give the squared error loss from the heuristic methods we discussed. Column B uses the values for $(x_1, x_2, ..., x_{n-1})$ from the algorithm in Section 3.1 applied to the X distribution. Column C uses $(y_1, y_2, ..., y_{n-1})$ from using the Y distribution. Column D gives the squared error losses from the averages of Y partition limits used in columns B and C, while Column E is based on the average X partition limits in columns B and C. Note that the values in columns D and E are extremely close to those in A, suggesting that the heuristic methods used in columns D and E perform well.

Table 4: Y Normal and X Uniform (Squared Error Loss)

n	A	B	C	D	E
3	.3595	.3637	.3677	.3597	.3597
4	.2941	.2988	.3030	.2944	.2943
5	.2629	.2672	.2707	.2632	.2630
6	.2456	.2494	.2523	.2459	.2457

Once we have determined the partition limits for a fixed pair of distributions (F_Y, F_X), we can get the limits for any member of the scale family of F_Y and F_X by appropriate scaling. For example, let $(y_1, y_2...., y_n)$ and $(x_1, ..., x_n)$ be the partition limits when F_Y is standard normal but F_X is uniform on $(-1, 1)$. Then, if we have a new F_X that is uniform on $(-\sigma, \sigma)$, the new partition limits are $(y_1, ..., y_n)$ and $(\sigma x_1, ..., \sigma x_n)$. This result can be established using the same arguments as in the last section.

Further, we can find the value of σ (or the X distribution within the scale family) that minimizes (2). It can be shown, using arguments similar to the last section, that the optimum value of σ is given by

$$\sigma = \frac{\sum_{i=1}^{n} \frac{\left(\int_{F_X^{-1}(F_Y(y_{i-1}))}^{F_X^{-1}(F_Y(y_i))} x f_X(x)dx\right)\left(\int_{y_{i-1}}^{y_i} y f_Y(y)dy\right)}{F_Y(y_i)-F_Y(y_{i-1})}}{\int_{-\infty}^{\infty} x^2 f_X(x)dx}. \tag{14}$$

For instance if F_X is uniform $(-\sigma, \sigma)$ and Y is standard normal, the above expression gives that for $n = 3$ the optimal value for σ is 1.1461. Using this value instead of $\sigma = 1$ would reduce the expected squared error loss to .2888 from the value of .3595 in Column A of Table 4. This result has practical applications in situations in which the distribution of one of the two components is artificially created in order to have a large enough range of values to match the variation of the second component. This problem is further studied in the next section.

3.4 Optimal X−Distribution Under Squared Error Loss

Consider the situation where one of the two components is relatively inexpensive and its distribution is artificially created to have a sufficiently large range of values to match the variation for the distribution of the second component. The tappet example presented earlier is one such case. The manufacturing process that produces the tappets usually results in variation in tappet heights (X) that is substantially smaller than the variation in distance

Y from the bottom of the camshaft to the top of the valve. In order to produce tappets with heights spanning the range of values needed, tappets are manufactured with a number of different nominal (mean) heights, so that the resulting X distribution is in fact a mixture distribution. Since this X distribution is artificially created, we study the problem of optimal selection of this distribution.

Note that since X_i and Y_i are independent,

$$\sum_{i=1}^{n} \mathrm{E}(Y_i - X_i)^2 p_i = \sum_{i=1}^{n} [\mathrm{var}Y_i + \mathrm{var}X_i + (\mathrm{E}(Y_i) - E(X_i))^2] p_i.$$

So, for any given sets of partition limits and conditional means, we want $\mathrm{var}X_i$ to be as small as possible. Thus, the optimal X distribution will have $\mathrm{var}X_i = 0$, which can be accomplished by taking the X distribution to be discrete with $X = x_i^d$ when Y is in $(y_{i-1}, y_i]$. Let X^D denote this discrete random variable. Then we can write

$$\sum_{i=1}^{n} \mathrm{E}(Y_i - X_i)^2 p_i = \mathrm{E}((Y - X^D)^2) \geq \mathrm{E}(Y - \mathrm{E}(Y|X^D))^2$$

so the values $x_1^d, ..., x_n^d$ should be chosen to be equal to the conditional means $\mathrm{E}(Y_1), ..., \mathrm{E}(Y_n)$. The expected squared error then becomes

$$\begin{aligned} \mathrm{E}((Y - X^D)^2) &= \sum_{i=1}^{n} \mathrm{E}(Y_i - \mathrm{E}(Y_i))^2 p_i \\ &= \mathrm{E}(Y^2) - \sum_{i=1}^{n} (\mathrm{E}(Y_i))^2 p_i. \end{aligned} \tag{15}$$

Note that when X and Y have the same distribution, (4) is twice (15). Hence, the set of partition limits $y_1, ..., y_{n-1}$ minimizes (15) if and only if it minimizes (4).

Thus, the optimal Y partition limits derived in Section 3.1 are also optimal for the problem presented in this section. To find the partition limits and optimal X distribution to minimize squared error loss for a given Y distribution, one should use the optimal Y partition limits from Section 3.1 and choose X to be the discrete distribution taking the value $\mathrm{E}(Y_i)$ whenever Y is in $(y_{i-1}, y_i]$. While this result is only optimal in theory, (since forcing a continuous X to take only n discrete values may be impossible in practice), it does provide an approximately optimal strategy when the natural variation in the X components is small relative to the variation in Y such that the resulting X mixture distribution is practically discrete. Furthermore, if the X component is cheap relative to Y, it may be practical to further force discreteness in the X distribution by discarding any components which differ substantially from the desired set of discrete values. The current practice in many industries is to fabricate the cheaper product according to a uniform distribution within the bins. The above argument shows that this is clearly sub-optimal and that we should try to minimize the variation within the bins.

The equivalence of (4) and (15) also shows that, for squared error loss, the optimization problem for selective assembly is related to a "one-sided" optimal partitioning problem that has been discussed in the literature (Eubank, 1988). In this problem there is only one distribution of interest and the goal is to choose the best discrete distribution to approximate the given distribution. Eubank (1988) provides an excellent review of these problems and discusses connections to various applications including piecewise constant approximation of a given function. Tarpey and Flury (1996) discuss related results on self-consistent and principal points and other applications.

4 Optimal Schemes Under Absolute Error Loss

We now consider optimal partitions that minimize the expected absolute difference from the target given by (3). We will again restrict attention to the case where X and Y have identical distributions, as in Section 3.1.

As before, let f denote the density of X (and Y) and let F be the corresponding cumulative distribution function. We will continue to require p_i^X is equal to p_i^Y for all i so we have $(y_1, y_2, ..., y_{n-1}) = (x_1, x_2, ..., x_{n-1})$. Thus we can rewrite (3) as

$$\sum_{i=1}^{n} \frac{\int_{x_{i-1}}^{x_i} \int_{x_{i-1}}^{x_i} |y-x| f(x) f(y) dx dy}{F(x_i) - F(x_{i-1})} =$$

$$2 \sum_{i=1}^{n} \frac{\int_{x_{i-1}}^{x_i} \int_{x_{i-1}}^{y} (y-x) f(x) f(y) dx dy}{F(x_i) - F(x_{i-1})}. \tag{16}$$

Taking the derivative with respect to each x_i $(1 \leq i \leq n-1)$ and setting them equal to zero shows that the optimal partition limits must satisfy the equations

$$x_i = \frac{\mathrm{E}(X_i) + \mathrm{E}(X_{i+1})}{2} + \frac{w_i - w_{i+1}}{4}, 1 \leq i \leq n-1. \tag{17}$$

where the w_i's are given by

$$w_i \equiv \mathrm{E}|Y_i - X_i| = \frac{\int_{x_{i-1}}^{x_i} \int_{x_{i-1}}^{x_i} |y-x| f(x) f(y) dx dy}{(F(x_i) - F(x_{i-1}))^2}.$$

Note the structure of this expression in comparison to (8) for the squared error loss case. The partition limits for the squared error loss are now adjusted by the w_i's, the expected value of the absolute differences. As in Section 3.1, equation (17) does not yield a closed form expression for the optimal partitions, but we can use an appropriate modification of the fixed-point algorithm in Section 3.1 to compute the limits for any given distribution.

Table 5 gives the resulting optimal partition limits for the standard normal distribution truncated at ± 3 along with comparisons to the methods of equal width and equal area partitioning in terms of expected absolute difference from the target as given by (3). Note that the (non-negative) partition limits are smaller than the corresponding ones for squared error loss in this example. Also, the percentage savings over the equal area and equal width methods are not as substantial as with squared error loss.

Table 5: Standard Normal Distribution Truncated at -3 and 3 (Absolute Error Loss)

n	Optimal Partition Limits (Only the nonnegative values are given since all partitions are symmetric.)							Expected Absolute Difference from Target	Percentage Savings Over Equal Area Partitioning	Percentage Savings Over Equal Width Partitioning
1	-							1.117	-	-
2	0.000							.651	0%	0%
3	0.526							.462	1.24%	18.89%
4	0.000	0.838						.358	2.56%	20.83%
5	0.323	1.056						.292	3.78%	21.92%
6	0.000	0.554	1.222					.247	4.87%	22.36%
7	0.234	0.732	1.354					.214	5.83%	22.53%
8	0.000	0.417	0.876	1.464				.188	6.70%	22.58%
9	0.184	0.566	0.997	1.557				.168	7.47%	22.57%
10	0.000	0.335	0.691	1.099	1.637			.152	8.16%	22.54%
11	0.152	0.463	0.798	1.188	1.707			.139	8.80%	22.50%
12	0.000	0.280	0.573	0.892	1.267	1.769		.128	9.37%	22.45%
13	0.129	0.392	0.669	0.975	1.337	1.825		.118	9.89%	22.41%
14	0.000	0.241	0.489	0.754	1.049	1.400	1.875	.110	10.38%	22.36%
15	0.112	0.340	0.576	0.831	1.116	1.457	1.921	.103	10.82%	22.32%

5 Nonequal Partition Probabilities

When only one of the components is binned and the cost of that binned component is relatively cheap, the restriction that the p_i^x must equal the p_i^y is no longer compelling. We explore the consequences of relaxing this assumption in this section. We restrict attention to squared error loss.

Consider the expected squared difference from the target given by (2). The average proportion of assemblies that will be made in which the two components come from the ith partition of their respective dimensional distributions is now given by p_i^Y. We can now rewrite (2) as

$$\sum_{i=1}^{n} \mathrm{E}(Y_i - X_i)^2 p_i^Y. \tag{18}$$

We can still determine one set of partition limits given the other in this section, but the second set will be chosen to minimize (18) rather than to maintain equal partition probabilities as in previous sections. To see how this is done, let us first fix the X partition limits $(x_1, x_2, ..., x_{n-1})$. Once a measurement of Y is taken, the index of the bin from which X will selected will be a function of that Y value. We can express this by writing (18) as

$$\mathrm{E}(Y - X_{\delta(Y)})^2 \tag{19}$$

where the function $\delta(y)$ is equal to i when $y \in (y_{i-1}, y_i]$. Now the function δ will minimize (19) provided it minimizes

$$\mathrm{E}[(Y - X_{\delta(Y)})^2 | Y = y] = (y - \mathrm{E}(X_{\delta(y)}))^2 + \mathrm{var}(X_{\delta(y)})$$

for all values of y. Note that

$$(y - \mathrm{E}(X_i))^2 + \mathrm{var}(X_i) = (y - \mathrm{E}(X_{i+1}))^2 + \mathrm{var}(X_{i+1})$$

when y is equal to

$$\frac{\mathrm{E}(X_i) + \mathrm{E}(X_{i+1})}{2} + \frac{\mathrm{var}(X_i) - \mathrm{var}(X_{i+1})}{2\mathrm{E}(X_i) - 2\mathrm{E}(X_{i+1})}. \tag{20}$$

It follows that if the values of (20) form an increasing sequence in i, then they minimize (18) for a given set of X partition limits $(x_1, x_1, ..., x_{n-1})$.

Although (20) gives a closed form expression for $(y_1, y_2, ..., y_{n-1})$ as a function of $(x_1, x_1, ..., x_{n-1})$, the functional form is quite complicated and thus differentiation of (18) with respect to the x_i does not lead to useful algorithms for finding the optimal partitions. In practice, we need to carry out a numerical search over different values for the x_i, using (20) to find the corresponding y_i, and evaluating (18) to determine approximately where the minimum value is obtained. Table 6 shows the results of this when both X and Y have standard normal distributions.

Table 6: Standard Normal Distribution with Nonequal Partition Probabilities

n	Optimal X Partition Limits (Only nonnegative values are shown.)	Optimal Y Partition Limits (Only nonnegative values are shown.)	$\sum_{i=1}^{n} \mathrm{E}(Y_i - X_i)^2 p_i^Y$	Percentage Savings Over Equal X and Y Partition Probabilities
1	-	-	2.000	-
2	0.00	0.00	.727	0%
3	0.521	0.655	.375	1.51%
4	0.000 0.916	0.000 1.011	.232	1.24%
5	0.401 1.166	0.369 1.276	.158	1.34%
6	0.000 0.663 1.374	0.000 0.651 1.475	.115	1.22%

Note that the percentage savings in the squared error loss by allowing nonequal partition probabilities is negligible in this example. However, there are cases in which the savings can be substantial. For instance, if X and Y have significantly different scale parameters, then the optimal partitioning (and consequently the expected squared error loss) is largely influenced by whether or not one imposes equal partition probabilities. For example, if X is normal with mean zero and a standard deviation of 2 while Y is normal with mean of zero and a standard deviation of 1, the optimal X and Y partition limits in the case $n = 3$ are (-1.224,0,1.224) and (-.612,0,.612) leading to a value of 1.761 for (18). If nonequal partition probabilities are allowed, the optimal X and Y partitions limits are $(-.177, 0, .177)$ and $(-1.255, 0, 1.255)$ respectively, giving a value of .668 for (18), which is a reduction of greater than 60%. Of course, along with this reduction comes a significant waste of X components, as the upper and lower X partitions have a total probability of .929 but are only mated with 20.9% of the Y distribution.

A significant reduction in (18) can also be obtained if the binning of X is initially far from optimal and cannot be changed. For example, if X and Y each have standard normal distributions truncated at -3 and 3 and the X partition limits are fixed at $(-1, 0, 1)$ (i.e. equal width partitioning), then the optimal Y partition limits are given by (20) to be $(-.732, 0, .732)$. These Y partition limits result in an 11.1% reduction in (18) compared to using equal partition probabilities.

6 Camshaft Example Revisited

To illustrate the advantages of using optimal binnning strategies, we return to the example involving the camshaft, valve and tappet. A random assembly of camshafts and tappets was resulting in unacceptable variation, so the company involved in manufacturing this product switched to selective assembly with $n = 36$ bins. Equal width binning schemes are currently used to group the tappets. The target clearance is 300μm. The tappet distribution as well as the distribution for the gap between the camshaft and the top of the valve are approximated by truncated normal distributions. Based on available data, the parameters of these distributions (prior to truncation) are standard deviations of approximately 32.7μm and means of approximately 3200μm and 3500μm respectively. Since it is required that the partition probabilities are equal, the optimal bins to minimize the variance of the clearance (squared error loss) can be obtained using the method described in Section 3.1. The current equal width partitioning method yields a clearance variance of 4.939μm^2. The optimal squared error loss scheme will reduce this to 3.453μm^2 which is a 30% reduction over equal-width binning.

7 Discussion

Selective assembly is a cost-effective method for reducing variation in the overall assembly when component variation is unacceptably large. The overall variation decreases as the number of bins increases, but there are costs associated with using a larger number of bins, such as cost of taking more precise measurements and cost of keeping larger inventories to stock all bins. Thus it is desirable to use binning strategies that provide the greatest reduction in variation for any fixed number of bins. In this paper, we have studied such optimal binning strategies under both squared error and absolute error loss. For the specific distributions studied, these optimal strategies were shown to produce significant decreases in expected loss relative to the heuristic methods of equal width and equal area binning.

In some applications, asymmetric loss functions will be more appropriate. We have also not considered cases where the optimal bins should be chosen subject to an overall tolerance constraint. Another practical issue of interest is the presence of measurement in error in one or more of the component dimensions. These problems will be studied in future work.

Finally, it is important to note that we have not addressed the issue of limited buffer capacity; that is if there are no components of one type in a particular bin at a particular time, the mating components in the corresponding bin cannot be used. A number of papers in the literature (Kannan and Jayabalan (2001), Thesen and Jantayavichit (1999), and Fang and Zhang (1995)) have dealt with this issue.

Acknowledgments

D. Mease's and V. Nair's research was supported in part by NSF grants DMS-980328 and DMS-0204247 and by a Ford URP grant. The authors are grateful to Kerby Shedden for helpful comments.

Appendix

Proof of Proposition 1: Suppose there exist two solutions to (8) given by $(x_1^A, x_2^A, ..., x_{n-1}^A)$ and $(x_1^B, x_2^B, ..., x_{n-1}^B)$. Let the corresponding sets of conditional means corresponding to these two sets of class partitions be denoted by $\mathrm{E}(X_1)^A, \mathrm{E}(X_2)^A, ..., \mathrm{E}(X_n)^A$ and $\mathrm{E}(X_1)^B, \mathrm{E}(X_2)^B, ..., \mathrm{E}(X_n)^B$ respectively. Now let i^* be the largest i such $x_i^A \neq x_i^B$. Without loss of generality we will take $x_{i^*}^A < x_{i^*}^B$. Since (9) is non-increasing in t we have

$$\mathrm{E}(X_{i^*+1})^B - x_{i^*}^B \leq \mathrm{E}(X_{i^*+1})^A - x_{i^*}^A.$$

Now from (8) we have

$$\mathrm{E}(X_{i^*+1})^B - x_{i^*}^B = x_{i^*}^B - \mathrm{E}(X_{i^*})^B$$

and

$$\mathrm{E}(X_{i^*+1})^A - x_{i^*}^A = x_{i^*}^A - \mathrm{E}(X_{i^*})^A$$

so that

$$x_{i^*}^B - \mathrm{E}(X_{i^*})^B \leq x_{i^*}^A - \mathrm{E}(X_{i^*})^A.$$

Again using the fact that (9) is non-increasing in t it follows that

$$x_{i^*-1}^A < x_{i^*-1}^B$$

and

$$\mathrm{E}(X_{i^*})^B - x_{i^*-1}^B \leq \mathrm{E}(X_{i^*})^A - x_{i^*-1}^A$$

Now we can repeat the entire argument with i^* replaced by $i^* - 1$ to obtain

$$x_{i^*-2}^A < x_{i^*-2}^B$$

and

$$\mathrm{E}(X_{i^*-1})^B - x_{i^*-2}^B \leq \mathrm{E}(X_{i^*-1})^A - x_{i^*-2}^A.$$

Continuing inductively, we will eventually have

$$x_1^A < x_1^B$$

and

$$\mathrm{E}(X_2)^B - x_1^B \le \mathrm{E}(X_2)^A - x_1^A.$$

Finally, (8) implies

$$\mathrm{E}(X_2)^B - x_1^B = x_1^B - \mathrm{E}(X_1)^B$$

and

$$\mathrm{E}(X_2)^A - x_1^A = x_1^A - \mathrm{E}(X_1)^A$$

so that

$$x_1^B - \mathrm{E}(X_1)^B \le x_1^A - \mathrm{E}(X_1)^A.$$

Furthermore, it can be verified that the inequality must become strict eventually so that

$$x_1^B - \mathrm{E}(X_1)^B < x_1^A - \mathrm{E}(X_1)^A.$$

However, this last statement along with

$$x_1^A < x_1^B$$

contradicts the assumption that (9) is non-increasing in t when $h = -\infty$.

Proof of Lemma 1: Given $-\infty < t_1 < t_2 < \infty$ let the random variables W_1 and W_2 have density functions

$$\frac{f(t_1 + w)}{F(t_1 + h) - F(t_1)}$$

and

$$\frac{f(t_2 + w)}{F(t_2 + h) - F(t_2)}$$

respectively for $0 \le w \le h$ and zero otherwise and assume f is strongly unimodal. The result will follow provided we can show $\mathrm{E}[W_1] \ge \mathrm{E}[W_2]$, which is implied if the hazard rate for W_1 is less than or equal to that of W_2. Since these hazard rates are given by

$$\frac{f(t_1 + w)}{F(t_1 + h) - F(t_1 + w)}$$

and

$$\frac{f(t_2 + w)}{F(t_2 + h) - F(t_2 + w)}$$

respectively, it is sufficient to show

$$\frac{f(x)}{F(x + b) - F(x)}$$

is non-decreasing in x for all $b > 0$. Taking the derivative with respect to x gives a fraction with a squared denominator and numerator given by

$$f'(x)(F(x+b) - F(x)) - f(x)(f(x+b) - f(x))$$

(the derivative $f'(x)$ exists almost everywhere since f is strongly unimodal). The above expression must be non-negative for $b > 0$ since its limit is zero as b goes to zero from the right, and its derivative with respect to b is non-negative for $b > 0$ because

$$\frac{f'(x)}{f(x)} \geq \frac{f'(x+b)}{f(x+b)}$$

by the assumption that f is strongly unimodal. Thus (9) is non-increasing in t for all h in $(0, \infty)$.

References

Ceglarek, D. and Shi, J. (1995), "Dimensional Variation Reduction for Automotive Body Assembly," *Journal of Manufacturing Review*, 8, 139-154.

Eubank, R. L. (1988), "Optimal Grouping, Spacing, Stratification, and Piecewise Constant Approximation," *SIAM Review*, 30, 404-420.

Fang, X.D. and Zhang, Y. (1995), "A New Algorithm for Minimizing the Surplus Parts in Selective Assembly," *Computers and Industrial Engineering*, 28, 341-350.

Kannan, SM. and Jayabalan, V. (2001), "A New Grouping Method for Minimizing the Surplus Parts in Selective Assembly," *Quality Engineering*, 14, 67-75.

Kwon, H.M., Kim, K.J., and Chandra, M.J. (1999), "Economic Selective Assembly Procedure," *Naval Research Logistics*, 46, 809-821.

Pugh, G.A. (1986), "Partitioning for Selective Assembly," *Computers and Industrial Engineering*, 11, 175-179.

Pugh, G.A. (1992), "Selective Assembly with Components of Dissimilar Variance," *Computers and Industrial Engineering*, 23, 487-491.

Tarpey, T. (1994), "Two Principal Points of Symmetric, Strongly Unimodal Distributions," *Statistics and Probability Letters*, 20, 253-257.

Tarpey, T. and Flury, B. (1996), "Self-Consistency: A Fundamental Concept in Statistics," *Statistical Science*, 11, 229-243.

Thesen, A. and Jantayavichit, A. (1999), "Design and Evaluation of a Selective Assembly Station for High Precision Scroll Compressor Shells," *Proceedings of the 1999 Winter Simulation Conference*, 694-700.

Trushkin, A. (1982), "Sufficient Conditions for Uniqueness of a Locally Optimal Quantizer for a Class of Convex Error Weighting Functions," *IEEE Transactions on Information Theory*, 28, 187-198.