CHAPTER 2. THE RANDOMIZED RESPONSE TECHNIQUE

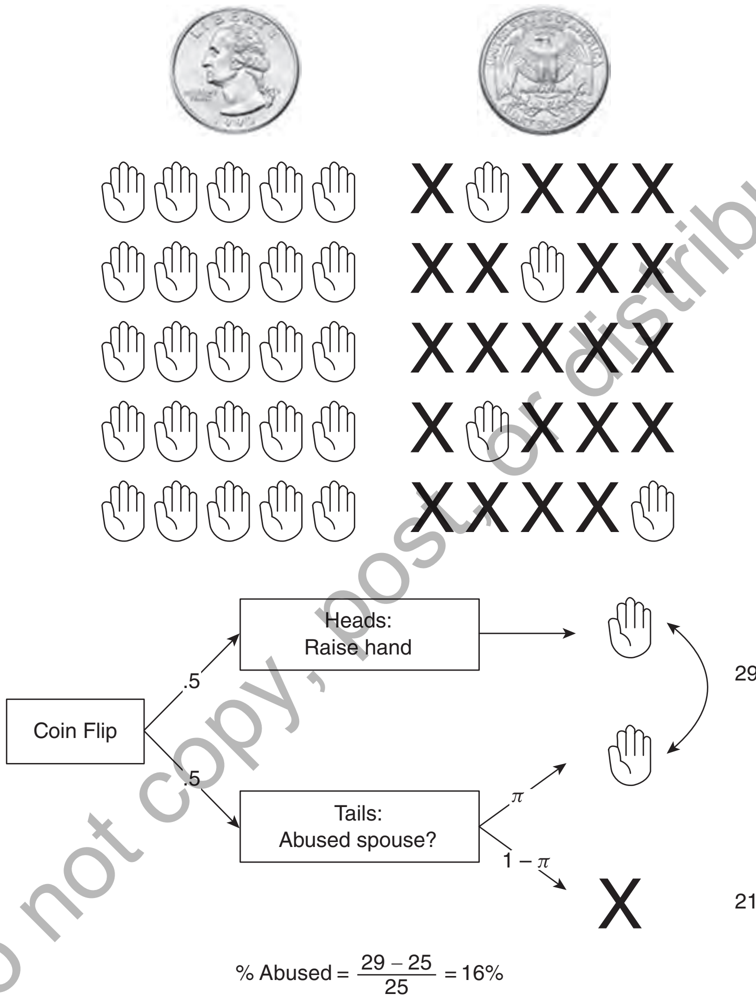
It may not be widely known that in an execution by firing squad, not all of the marksmen have guns supplied with live ammunition. In Utah, for example, the only state that has employed this form of capital punishment in recent decades, one of the five sharpshooters, at random, is given a rifle loaded with a blank. The purpose of this arrangement is so that each squad member can maintain that he had fired the blank, should his conscience force him to do so. Although it is possible after firing to discern on the basis of the rifle's recoil whether one has shot a real bullet or a blank, it is not known until after the trigger is pulled. One can rationalize, therefore, that the act of pulling the trigger was not necessarily fatal.[1]

The same use of probability to assuage the conscience of members of a firing squad can be applied in social surveys of sensitive information. Since survey respondents tend to conceal embarrassing or threatening information, particularly when an interviewer (in person or over the phone) might disapprove, the element of chance can be used to inoculate their responses yet at the same time provide the researcher with sufficient data for statistical analysis of prevalence rates, means, and variances, as well as estimating various measures of association.

To illustrate hypothetically, suppose that we wished to estimate the prevalence of spouse abuse by surveying a gathering of men in an auditorium.[2] Simply asking the men to raise their hands if they had ever beaten their spouse would most likely produce a roomful of rubbernecks, but few would be apt to volunteer an affirmative response through a raised hand. Instead, we could instruct each of the men to flip a coin privately and then to raise his hand *either* if he had ever abused his spouse or had obtained a head on the coin. Clearly, those raising their hands would hardly be implicated in abusive behavior, since most would be doing so because of getting a head on the coin flip. Even though we could not determine the true status of any of the hand-raisers, an estimate of abuse prevalence is available nonetheless.

Continuing with this hypothetical case, suppose that the room contained 50 men, 29 of whom had raised their hands. As shown in the probability tree at the bottom of Figure 2.1, we can expect that half ($p = .5$) of the men (i.e., 25) would have received heads on their coins and thus raised their hands, regardless of their true status on the abuse question. The excess of 4 hand-raisers would suggest that 4 of the 25 who expectedly received a tail on the flip had in fact abused their spouses. Therefore, we could estimate, based solely on the presumed (but unidentifiable) 25 tail recipients, the

Figure 2.1 Hypothetical Example of Randomized Response



prevalence of wife abuse to be 4/25 or 16%. Overall, we would expect there also to be 4 wife abusers among the 25 head recipients, all of whom raised their hands anyway. Overall, then, we would project there to be 8 abusers among the sample of 50 in the room, even though we could not determine which of the 29 hand-raisers constituted the 8 among them believed to have abused their wives.

This simple example only illustrates the logic underlying the randomized response approach. In practice, the range of estimation models developed over the past half-century is quite broad, and the variety of survey procedures for implementing randomized response for sensitive data collection is rich with innovation.

In the chapters to follow, we will highlight several variants of randomized response, some straightforward and others rather complicated, and outline how univariate and multivariate statistics can be derived from the data. Next, we will present the results of a number of methodological experiments comparing randomized response with alternative data collection strategies, as well as an array of practical applications of the method (and related approaches) for measuring sensitive topics of interest to social scientists.

Warner's Randomized Response Design

The similarity between not knowing which of 5 sharpshooters in a firing squad are the 4 to be supplied live bullets and which of the 29 hand-raisers in the hypothetical gathering of married men are the 8 (the 4 with tails plus a complementary 4 with heads) who had abused their spouse should be clear. In practice, one most likely would not attempt to survey large groups of respondents in convened settings like this if the subject matter is anything more sensitive than their favorite sports team. (If one could in fact gather one's sample together like this, in certain cases, an anonymous questionnaire may be preferred.) Still, the same procedure and the same level of protection can be applied in a one-on-one setting (in person or by phone).

This was the brainstorm of Stanley Warner—that the element of chance could inoculate responses to sensitive inquiries. His 1965 article, "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," revolutionized the way that the sample survey of sensitive topics could be designed. Although the relative crudeness of Warner's pioneering formulation limited the appeal of the randomized response technique initially, many improvements and enhancements, which we shall review here, have significantly increased the potential value of this method. Warner's seminal approach may not eliminate evasive answer bias, as the title of his article boldly claims; however, it clearly points out a probabilistic logic to sensitive surveys useful for reducing response bias.

In a survey of abortion (particularly in the days when abortions were strictly illegal), a woman who had terminated her pregnancy might falsely deny the statement, "I have had an abortion," or, conversely, falsely affirm the statement, "I have never had an abortion." That is, it is not necessarily the "True" or "False" answer that is stigmatizing or embarrassing, but the connection between the question and the response. Warner's suggestion was that a respondent be posed both questions but answer one or the other depending on the outcome of a randomizing device (e.g., the spinner illustrated in Figure 2.2) that only she sees. The response is no longer revealing, since no one except the respondent is aware of the question.

More formally, Warner's strategy directs the respondent to react to one of two logical opposites, depending on the outcome of a randomizing device, following a Bernoulli distribution with parameter p:

Statement 1: I am a member of A

Statement 2: I am not a member of A

where A represents some sensitive attribute (e.g., having had an abortion).

Figure 2.3 illustrates the Warner model with a probability tree diagram. As shown, p is the probability that a respondent is directed to answer the sensitive question and $1 - p$ is the probability that he or she is instructed to answer its converse. The respective probabilities of the presence and absence of the sensitive trait are then π and $1 - \pi$.

Elementary probability theory dictates that the total proportion (regardless of question) of affirmative responses, λ, can be expressed in terms of π, the probability of possessing the attribute in question (e.g., having had an abortion), in the following way:

$$\begin{aligned} \text{P(''True'' Response)} &= \text{P(Question 1)P(}A\text{ is True)} \\ &\quad + \text{P(Question 2)P(}A\text{ is False)} \end{aligned}$$

or more formally,

$$\lambda = p\pi + (1 - p)\,(1 - \pi)$$

Figure 2.2 Warner-Type Spinner

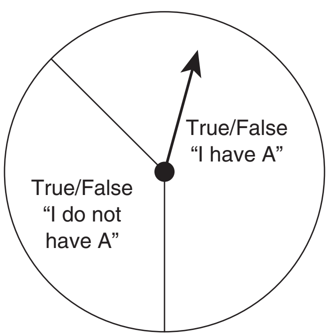
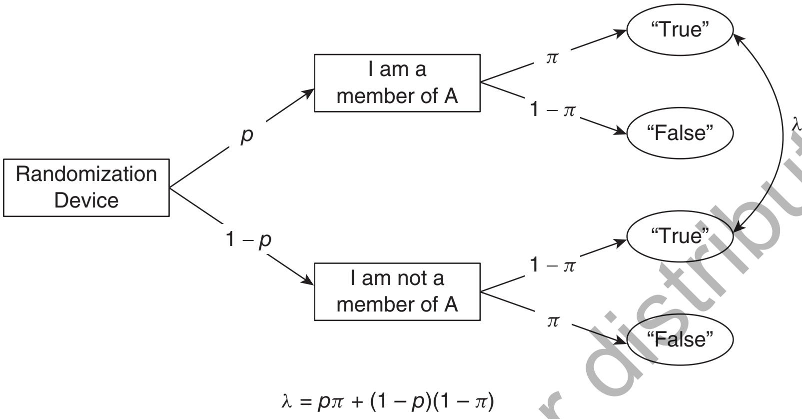


Figure 2.3 Probability Tree Diagram of the Warner Model



Solving for an estimate of π in terms of known p and $\hat{\lambda}$, the observed sample proportion answering "True,"

$$\hat{\pi} = \frac{\hat{\lambda} + p - 1}{2p - 1} \tag{1}$$

so long as $p \neq .5$. This estimate has a sampling variance,

$$\mathrm{Var}(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \tag{2}$$

As is the case with all measures of sampling variance to be presented throughout these pages, $\mathrm{Var}(\hat{\pi})$ can be estimated from the sample data, replacing n with $n - 1$ to incorporate the usual correction for bias. After square rooting, we obtain a standard error (SE) estimate,

$$SE(\hat{\pi}) = \hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n-1} + \frac{p(1-p)}{(n-1)(2p-1)^2}}$$

which can then be used to create a confidence interval for π. In addition, while all of the formulas presented throughout this monograph assume simple random sampling with replacement, a correction for a finite population may be necessary if sampling without replacement from a population that is not significantly larger in size than the sample drawn from it. For the Warner design, according to Kim and Flueck (1978), a finite population correction should be applied to the first term of the variance expression when sampling without replacement (WOR). Specifically,

$$\mathrm{Var}(\hat{\pi})_{\mathrm{WOR}} = \frac{\pi(1-\pi)}{n}\left(\frac{N-n}{N-1}\right) + \frac{p(1-p)}{n(2p-1)^2}$$

As an illustration of randomized response in practice, the IIT Research Institute (1971) surveyed the prevalence of various activities associated with organized crime in Illinois using the Warner design. Each respondent was instructed to draw at random from a container of 10 marbles, 2 of which were blue and 8 of which were green ($p = .2$) and then to answer affirmatively or negatively to the appropriate statement:

Blue marble:	I have used heroin
Green marble:	I have never used heroin

This approach was repeated for a range of illegal activities, mainly involving drugs and gambling.

Despite the innovation, Warner's method had two severe limitations. First, as can be seen from Equation (2), the variance (and thus the standard error) of the estimator is considerably inflated over what it would be if respondents were asked the sensitive question directly. Specifically, the first term of the equation is the usual variance of a sample proportion, and the second term, therefore, represents the additional sampling error due to the randomizing procedure.

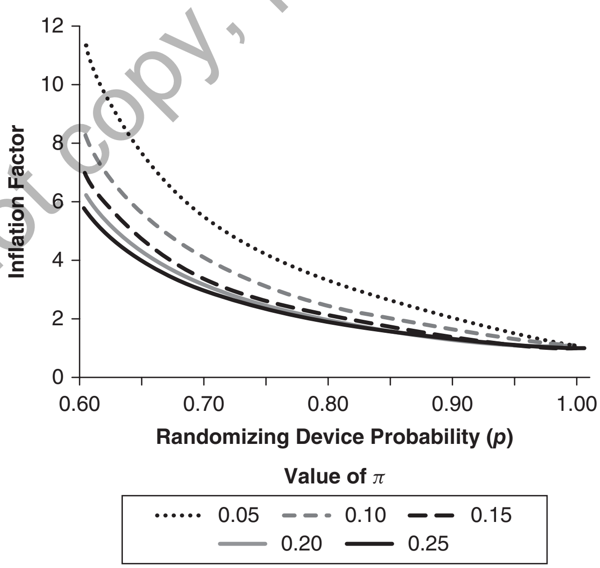
Figure 2.4 displays the impact that p, the probability of question selection, has on the estimation process, specifically on the inflation in standard error over conventional direct questioning, for several alternative values of π, the prevalence of the attribute A. The closer that p is to .5 (and thus the more ambiguous the response), the greater the inflation in standard error when estimating π. Suppose, for example, that a die is used, and the respondent is told to consider the affirmative statement "I am a member of A" if the roll shows 1 through 4, and the negative statement "I am not a member of A" otherwise (i.e., $p = 4/6 = .67$). If $\pi = .20$, then the standard error is nearly quadrupled. An even larger choice of p of $5/6 = .833$ causes the standard error and thus the confidence interval for estimating π to be doubled in size. Moreover, it is mathematically possible, particularly with a small sample size and a selection probability p closer to .5, to operate under standard error $SE(\hat{\pi})$ that is so inflated that prevalence estimates >1 or <0 can result (e.g., see Brown, 1975; Reaser, Hartsock, & Hoehn, 1975). However, as Lee, Sedory, and Singh (2013b) have demonstrated through a set of simulations for various values of p and π, even modest sample sizes of a couple hundred or more can virtually ensure that such nonsensical estimates will not occur in practice.

The second concern with Warner's model is less statistical in nature. Under this scheme, both questions deal with a very sensitive topic. Some respondents may wonder if there is a mathematical trick that will permit the interviewer to figure out what their true status is. Even if the respondent appreciates the protection offered by the approach, the threatening nature of both questions may encourage a refusal to participate.

Unrelated Question Approach

In an attempt to desensitize the use of an embarrassing question paired with its converse, Simmons proposed that the sensitive question be paired in a similar randomized procedure with an innocuous question (see Horvitz,

Figure 2.4 Standard Error Inflation With the Warner Model



Shah, & Simmons, 1967). The so-called unrelated question approach might present these items:

Question 1: Have you ever had an abortion?

Question 2: Did you go to the movies within the past month?

In this design, two parameters are unknown: π_x and π_y, the prevalence of abortion and of movie attendance, respectively. Employing (necessarily) two independent samples with different selection probabilities p_1 and p_2 ($p_1 \neq p_2$), the respective probabilities of affirmative responses in the two samples are given by the following:

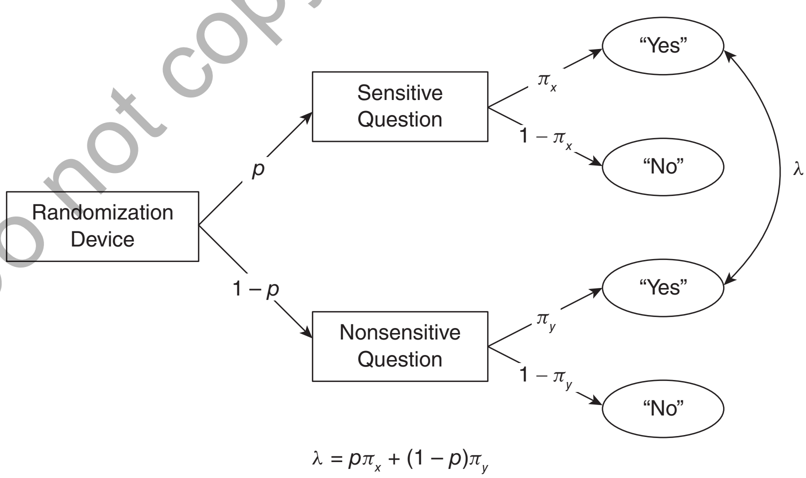
$$\lambda_j = p_j \pi_x + (1 - p_j)\pi_y$$

for $j = 1, 2$. Figure 2.5 illustrates the unrelated question model using a probability tree diagram.

The desired estimate of π_x in terms of sample proportions of "Yes" responses, $\hat{\lambda}_j$, becomes

$$\hat{\pi}_x = \frac{\hat{\lambda}_1(1 - p_2) - \hat{\lambda}_2(1 - p_1)}{p_1 - p_2} \tag{3}$$

Figure 2.5 Probability Tree Diagram of the Unrelated Question Model



with sampling variance

$$\operatorname{Var}(\hat{\pi}_x) = \frac{1}{(p_1 - p_2)^2}\left(\frac{\lambda_1(1-\lambda_1)\,(1-p_2)^2}{n_1} + \frac{\lambda_2(1-\lambda_2)(1-p_1)^2}{n_2}\right) \tag{4}$$

In a study of drug use among a sample of Ontario high school students, for example, Goodstadt and Gruson (1975) utilized this two-group, two unrelated questions design with the last digit of the subject's telephone number serving as the randomizing device. The students were significantly more likely to respond to drug use questions under the randomized response condition than under traditional direct questioning. Estimates pertaining to the use of six drugs during the preceding 3 months also were significantly higher for randomized response than for direct questions for five of the drugs, including (1) alcohol, (2) cannabis, (3) amphetamines, (4) tranquilizers, and (5) heroin. Only in the case of hallucinogens were the estimated usage rates essentially the same for the two survey approaches.

One can greatly simplify the unrelated question design and eliminate the need for splitting the sample in two by using an alternative response whose distribution is known in advance. For example, the question, "Were you born in November?" gives $\pi_y = 30/365$.[3] As a result, Equations (3) and (4) reduce to

$$\hat{\pi}_x = \frac{\hat{\lambda} - (1-p)\pi_y}{p} \tag{5}$$

and

$$\operatorname{Var}(\hat{\pi}_x) = \frac{\lambda(1-\lambda)}{np^2} \tag{6}$$

Moors (1971) noted that, even when questions having known distributions are not readily available, the simplicity of the known π_y model can be achieved if one of the two samples in the Simmons unrelated question model were used exclusively to estimate an unknown π_y (i.e., $p_2 = 0$). In other words, a sensitive question is paired with a nonsensitive alternative in the first sample, and the nonsensitive question is asked directly in the second sample. Moors demonstrated that this modification could in fact produce a more efficient estimate of the desired parameter π_x.[4] Specifically, by setting $p_2 = 0$, which allows $\hat{\lambda}_2$ to be a direct estimate of π_2, Equations (3) and (4) simplify to the following:

$$\hat{\pi}_x = \frac{\hat{\lambda}_1 - (1 - p_1)\hat{\lambda}_2}{p_1} \tag{7}$$

and

$$\mathrm{Var}(\hat{\pi}_x) = \frac{1}{p_1^{\,2}}\left(\frac{\lambda_1(1-\lambda_1)}{n_1} + \frac{\lambda_2(1-\lambda_2)(1-p_1)^2}{n_2}\right) \tag{8}$$

Taking Moors's improvement one step further, Folsom, Greenberg, Horvitz, and Abernathy (1973) recognized that the two samples could be exploited far more efficiently by using two alternative questions, and not "wasting" one entire sample on just estimating the instrumental parameter π_y. As in Moors's plan, Folsom and his colleagues suggested that the nonsensitive parameters be estimated directly. As in Simmons's model, both samples are still used for estimating the desired sensitive parameter. Specifically, in the first sample, one of the nonsensitive questions is paired with the sensitive inquiry and the other nonsensitive question is posed directly. In the second sample, the roles of the two nonsensitive questions are reversed.

For example, in the first sample, respondents could be instructed to flip a coin and respond to a pair of questions as directed and then answer an additional question outright:

HEAD:	Did you cheat on your income taxes last year?
TAIL:	Did you watch the 11:00 news last night?
DIRECT:	Do you have any older living siblings?

Respondents in the other sample would then follow the same procedure but with a swap between the two nonsensitive questions:

HEAD:	Did you cheat on your income taxes last year?
TAIL:	Do you have any older living siblings?
DIRECT:	Did you watch the 11:00 news last night?

For the first sample,

$$\lambda_1^r = p\pi_x + (1-p)\pi_{y_1}$$
$$\lambda_1^d = \pi_{y_2}$$

and for the second sample,

$$\lambda_2^r = p\pi_x + (1-p)\pi_{y_2}$$
$$\lambda_2^d = \pi_{y_1}$$

where λ_j^r is the probability of an affirmative response to the randomized response pair in the jth sample; λ_j^d is the probability of an affirmative response to the direct question in the jth sample; and p is the probability of selecting the sensitive question in both samples. Because, in essence, Moors's model is applied twice, two unbiased estimates derive from the sample proportions of affirmative responses:

$$\hat{\pi}_x(1) = \frac{\hat{\lambda}_1^r - \hat{\lambda}_2^d(1-p)}{p}$$
$$\hat{\pi}_x(2) = \frac{\hat{\lambda}_2^r - \hat{\lambda}_1^d(1-p)}{p} \tag{9}$$

having respective variances

$$\mathrm{Var}[\hat{\pi}_x(1)] = \frac{1}{p^2}\left(\frac{\lambda_1^r(1-\lambda_1^r)}{n_1} + \frac{(1-p)^2\pi_{y_1}(1-\pi_{y_1})}{n_2}\right)$$
$$\mathrm{Var}[\hat{\pi}_x(2)] = \frac{1}{p^2}\left(\frac{\lambda_2^r(1-\lambda_2^r)}{n_2} + \frac{(1-p)^2\pi_{y_2}(1-\pi_{y_2})}{n_1}\right) \tag{10}$$

A final estimate of π_x is given by a weighted average

$$\hat{\pi}_x = w\hat{\pi}_x(1) + (1-w)\hat{\pi}_x(2)$$

where w is chosen to minimize the variance of the weighted average, which is achieved by setting w and $1 - w$ to be inversely proportional to $\mathrm{Var}[\hat{\pi}_x(1)]$ and $\mathrm{Var}[\hat{\pi}_x(2)]$. If the sample sizes are nearly equal, however, equal weights are usually sufficient.

Rather than having to split the sample into two groups, to select two non-sensitive questions for pairing with the sensitive item, and to weight two competing prevalence estimates for their relative sampling variances, Mangat (1994) offered a fairly straightforward approach for optimizing the efficiency of randomized response. According to his design, those who do and those who do not have the sensitive attribute are to follow different instructions for answering the survey. Respondents having the sensitive characteristic are prompted to answer the question truthfully, while the others are directed (in a Warner-type fashion) to use a randomized device (with probability $p \neq .5$) and then either to confirm falsely or deny truthfully the sensitive trait. Essentially, all denials are truthful, whereas only the affirmative responses are ambiguous by design. A probability tree diagram of Mangat's approach is displayed in Figure 2.6.

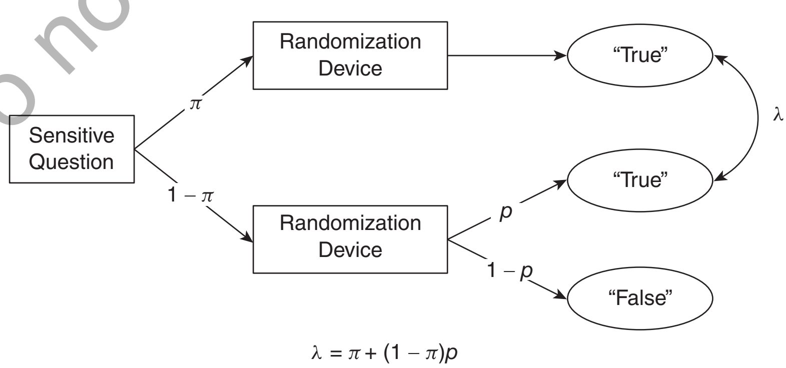
For example, in a survey of sexual infidelity, a sample of married men and women might be asked to roll a die and then respond with the following question pattern:

If cheated:	Say "True"
If did not cheat:	Say "True" if rolled a 1 or 2
	Say "False" if rolled a 3 through 6

If, as before, $\hat{\lambda}$ is the observed proportion of respondents indicating "True," then an estimate of the prevalence of the sensitive attribute (e.g., infidelity) would be

$$\hat{\pi} = \frac{(\hat{\lambda} - 1 + p)}{p}$$

Figure 2.6 Probability Tree Diagram of the Mangat Model



with sampling variance

$$\text{Var}(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(1-p)}{np}$$

The extreme sensitivity of certain behaviors or traits is sometimes related, at least in part, to their rarity. After all, the level of embarrassment associated with them might not be so profound were they more commonplace in the general population. Unfortunately, estimating the prevalence of rare events is challenging no matter what survey method is used. When studying sensitive attributes with exceptionally low prevalence π_x, the randomized response designs considered thus far can be modified to improve the quality of estimates by appealing to the Poisson approximation of the binomial so long as the sample size is quite large. With $\pi_x \to 0$ and $n \to \infty$, the number of individuals having the sensitive trait follows a Poisson distribution with parameter $\theta_x = n\pi_x$. The unrelated question randomized response approach can be used, for example, in pairing the sensitive question with a nonsensitive alternative that also addresses a rare characteristic, but with known probability π_y, and so $\pi_y \to 0$, $n \to \infty$, and the number with a true response follows a Poisson distribution with parameter $\theta_y = n\pi_y$.

For example, a respondent might be asked to say "Yes" or "No" to one of the following two questions with probabilities p and $1 - p$, respectively, where the selection of which question to answer depends on the outcome of the randomized device:

Question 1: Have you engaged in group sex within the past year?

Question 2: Were you born on January 1?

Denoting π_z as probability of an affirmative response to either question in the pair, if $\pi_z \to 0$ and $n \to \infty$, then the number of true responses obtained from the sample also has a Poisson distribution with parameter $\theta_z = n\pi_z$ as a mixture of the two underlying response distributions:

$$\theta_z = p\theta_x + (1 - p)\theta_y$$

where the probability of selecting the sensitive question p is accomplished by whatever means (spinner, dice, coin, etc.).

Letting $z_1, \ldots, z_n$ represent a series of ones and zeroes for affirmative and denial responses, respectively, Land, Singh, and Sedory (2012) have shown that a maximum likelihood estimate of the θ_x is given by

$$\hat{\theta}_x = \frac{\bar{Z} - (1-p)\theta_y}{p} \tag{11}$$

with sampling variance of

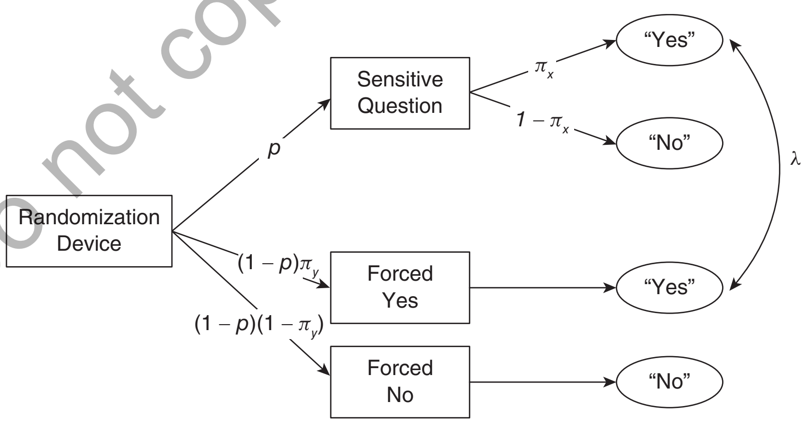
$$\mathrm{Var}(\hat{\theta}_x) = \frac{\theta_x}{np} + \frac{(1-p)\theta_y}{np^2} \tag{12}$$

The prevalence of the sensitive attribute can then simply be determined, $\hat{\pi}_x = \hat{\theta}_x / n$.

Forced Alternative Response

Nonsensitive alternative questions are employed in various randomized response models presented above to neutralize responses through question ambiguity. Actually, it is not necessary to have a nonsensitive question at all; rather, let the randomizing device itself generate the nonsensitive "responses." In the so-called *forced alternative* (or *forced response*) approach, various outcomes of the randomizing device have respondents give an artificial Yes/No response without there being a question or direct them to answer the sensitive question with a genuine Yes/No response. As shown in Figure 2.7, the randomizing device generates both p, the probability of a respondent being directed to answer the sensitive question, and π_y, the probability of a forced "Yes" response otherwise.

Figure 2.7 Probability Tree Diagram of the Forced Alternative Model



$\lambda = p\pi_x + (1 - p)\pi_y$

With a forced alternative design, for example, a sample of college students could be instructed to roll a pair of dice[5] and respond as follows:

Roll 2 through 4:	Just say "Yes"
Roll 5 through 9:	Have you engaged in sexual intercourse without a condom?
Roll 10 through 12:	Just say "No"

In this case, p = P(rolling 5–9) = .67, and π_y = P(rolling 2–4|rolling 2–4 or 10–12) = .5, and Equations (5) and (6) for an unrelated question with a known distribution can then be used. Note that the hypothetical illustration presented earlier in which men were asked to raise their hands either if they got a head on the coin or had abused their spouse is a forced alternative design with $p = .5$ and $\pi_y = 1.0$. In practice, however, requiring all forced alternative responses to be in the affirmative would give those for whom the sensitive trait does not apply no chance of providing a denial response, thereby encouraging noncompliance.

Although it may seem sterile and contrived to some respondents, the forced alternative approach has considerable appeal. First, the "non-sensitive" parameter, π_y, is fixed by design; not only is the model simplified with only π_x needing to be estimated, but the sampling variance of the estimate of this parameter is reduced. Second, since the forced responses can be a part of the randomizing device itself, the procedure may be easier for respondents to comprehend and follow, rather than confusing them with two questions.

Despite its simplicity, the forced response model may not sufficiently relieve respondents' privacy concerns. In a limited field test of the forced alternative method, Edgell, Himmelfarb, and Duchan (1982) observed considerable resistance among respondents to saying "Yes" when directed to do so, particularly when the sensitive behavior was having had a homosexual experience. Specifically, they used a microprocessor that would supply the random digits 0 through 9 on a screen. The digits 0 and 1 directed a "Yes" response, while 8 and 9 directed a "No" response. The remaining digits directed the respondent to answer a particular question. Among a series of 55 randomized inquiries, 13 questions had "random" digits that were, unbeknownst to the respondents, preprogrammed to fixed values. Three of the setup questions forced a "Yes" response (i.e., a 0 or 1 would appear for all respondents). For these, 15% reported "No" to a tax cheating question, despite the "Yes" direction; 11% refused to say "Yes" when tied to a question of whether they favored the construction of a nuclear power plant nearby; and 26% said "No" although directed otherwise on a question concerning homosexual experiences. On four setup items having a directed "No" in every case, nearly all of the respondents complied. The findings reported by Edgell and his colleagues are noteworthy, yet they were based on a small (n = 54) convenience sample of college students. Clearly, more research should focus on those conditions (e.g., type of sensitive question, or levels of p and π_y) that improve this form of randomized response.

Repeat Randomization

Rather than focusing on ways to pair sensitive questions with nonsensitive alternatives, Kuk (1990) offered a modification of randomized response, which he termed *indirect response*. The logic of Kuk's method is akin to that of Warner's original design, but it substitutes True/False responses to a sensitive question with some other binary response (e.g., color choices) that appears neutral even though it is linked probabilistically to the sensitive attribute under investigation. Moreover, Kuk's design utilizes repeated draws from a randomizing device.

To illustrate, one might create two decks of cards, each having a different mixture of red and blue circled card faces: Deck 1 with p_1 proportion of red cards and $(1 - p_1)$ proportion of blue cards, and Deck 2 with p_2 red cards and $(1 - p_2)$ proportion of blue cards ($p_1 \neq p_2$). The respondent is then instructed to draw one card at random from each of the two decks. Next, while privately looking at the pair of cards drawn, he or she is directed to indicate the color of the card selected from Deck 1 if the sensitive attribute applies and the color of the card drawn from Deck 2 if that attribute does not apply. Note that the color of the two drawn cards may or may not be the same. This procedure is illustrated with the probability tree diagram shown in Figure 2.8.

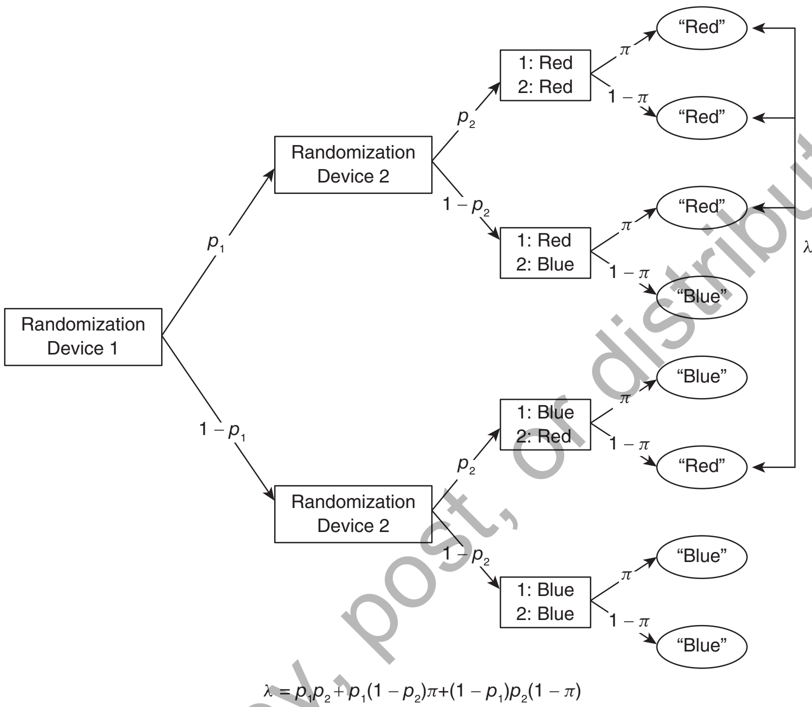
The expected proportion of respondents indicating "Red" as the card color from the random draw from the deck associated with their status on the sensitive question is given by

$$\lambda = p_1\pi + p_2\,(1 - \pi)$$

where π is the prevalence of the sensitive attribute. From this, we can estimate the prevalence of the sensitive trait by

$$\hat{\pi} = \frac{\hat{\lambda} - p_2}{p_1 - p_2} \qquad (13)$$

Figure 2.8 Probability Tree Diagram of the Kuk Model



as a function of the observed sample proportion of "Red" responses $\hat{\lambda}$, subject to sampling variance of

$$\mathrm{Var}(\hat{\pi}) = \frac{\lambda(1-\lambda)}{n(p_1 - p_2)^2} \tag{14}$$

In his paper, Kuk also advocated a more elaborate and statistically efficient extension involving repeated draws with replacement after each selection. Specifically, respondents would be instructed to make $k > 1$ selections from each deck (with replacement) and report the number of red cards obtained from the deck corresponding to their status on the sensitive attribute. Denoting $\hat{\lambda}_k$ as the proportion of all $k \times n$ draws reported as red cards, the prevalence estimate and its variance generalize to

$$\hat{\pi} = \frac{\hat{\lambda}_k - p_2}{p_1 - p_2} \tag{15}$$

and

$$\mathrm{Var}_k(\hat{\pi}) = \frac{\lambda(1-\lambda)}{kn(p_1 - p_2)^2} + \frac{\pi(1-\pi)}{n}\left(1 - \frac{1}{k}\right) \tag{16}$$

The larger k becomes (i.e., the greater the number of cards drawn per deck), the lower the variance of the estimate. However, asking respondents to make multiple draws with replacement, while keeping track of the number of cards matching a particular color, can be quite cumbersome and taxing for at least some of them. Respondent resistance and errors in keeping tabs on card counts would generally outweigh any advantage in estimation efficiency.

While using repeated draws from a randomizing card deck, it could be advantageous to elicit repeated responses as well. For example, as an efficiency boost to Warner's early design of pairing a statement with its complement, Odumade and Singh (2009) proposed a two-deck approach for estimating π in which a sample of n respondents is presented with the Warner-type arrangement twice, but with differing question selection probabilities. Specifically, a respondent is instructed to draw one card from each of the two decks containing mixtures of the two statements, "I am a member of A" and "I am not a member of A." After privately drawing the two cards, he or she is to indicate, in order of draw, whether the statements are true or false. Denoting p_1 and p_2 as the proportion of cards in Deck 1 and Deck 2 that are in the affirmative, the probabilities of the four possible response pairs (True/True, True/False, False/True, and False/False) are as follows:

$$\lambda_{\mathrm{TT}} = p_1 p_2 \pi + (1 - p_1)(1 - p_2)(1 - \pi)$$

$$\lambda_{\mathrm{TF}} = p_1 (1 - p_2)\, \pi + (1 - p_1)\, p_2 (1 - \pi)$$

$$\lambda_{\mathrm{FT}} = (1 - p_1)\, p_2 \pi + p_1 (1 - p_2)(1 - \pi)$$

$$\lambda_{\mathrm{FF}} = (1 - p_1)(1 - p_2)\, \pi + p_1 p_2 (1 - \pi)$$

After a fair amount of algebra and substituting the response probabilities with sample proportions, an estimate of π is provided by

$$\hat{\pi} = \frac{1}{2} + \frac{(p_1 + p_2 - 1)(\hat{\lambda}_{\mathrm{TT}} - \hat{\lambda}_{\mathrm{FF}}) + (p_1 - p_2)(\hat{\lambda}_{\mathrm{TF}} - \hat{\lambda}_{\mathrm{FT}})}{2[(p_1 + p_2 - 1)^2 + (p_1 - p_2)^2]} \tag{17}$$

with sampling variance

$$\mathrm{Var}(\hat{\pi}) = \frac{(p_1 + p_2 - 1)^2[p_1 p_2 + (1 - p_1)(1 - p_2)] + (p_1 - p_2)^2[p_2(1 - p_1) + p_1(1 - p_2)]}{4n[(p_1 + p_2 - 1)^2 + (p_1 - p_2)^2]^2} - \frac{(2\pi - 1)^2}{4n} \tag{18}$$

As another gesture toward greater efficiency, Singh and Grewal (2013) proposed a modification of Kuk's two-deck approach that involves properties of the geometric distribution. Each deck contains a mixture of cards marked "True" and "False," with the affirmative cards in proportions p_1 and p_2 ($p_1 \neq p_2$), respectively. Respondents are instructed to select cards from the first deck if the sensitive attribute is true for them and from the second deck if it is false. They are told to continue drawing cards with replacement from the appropriate deck until they obtain a card that bears their status with respect to the sensitive trait and only report the number of cards that were drawn.

The prevalence estimate is then

$$\hat{\pi}_x = \frac{p_1 p_2 \bar{Z} - p_1}{p_2 - p_1} \tag{19}$$

where $\bar{Z}$ is the average number of draws across all respondents. The estimate is subject to sampling variability

$$\mathrm{Var}(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{\pi p_2^2(1 - p_1) + (1 - \pi) p_1^2 (1 - p_2)}{n(p_2 - p_1)^2} \tag{20}$$

As with Kuk's own refinement of his basic design, Singh and Grewal's (2013) modification is also fairly demanding in terms of the steps required of respondents: choosing the appropriate deck, selecting cards with replacement after each draw, and accurately recounting the number of draws until a card is obtained that matches their status.[6] Despite the improved precision, the possible loss of cooperation may just be too much to risk.

Notes

1. A 20% chance of not firing a bullet may not always be sufficient protection for the shooters, thereby encouraging a marksman to shoot purposely off target so that another of the shooters will fire the fatal bullet. In the 1951 execution of Elisio J. Mares, Utah correctional officials were embarrassed when all four bullets missed the cloth target placed over the condemned man's heart; four bullets penetrated the right side of Mares's chest, killing him, although neither instantaneously nor painlessly (see Sifakis, 2003, p. 81).
2. For the purposes of this illustration, we shall ignore the obvious nonrepresentativeness of any such gathering, be it a lodge meeting or a Parent–Teacher Association assembly.
3. Although the distribution of birth month is not exactly uniform, the proportion of births in the month of November is virtually equal to this value.
4. Singh, Singh, and Mangat (2000) have noted a flaw in Moor's plan, although the likelihood that it will emerge in survey applications may be small. Under random sampling with replacement, there is the potential for a respondent to be randomly selected for both groups. As a result, the individual's direct response in Group 2 may expose the meaning of his or her response to the Group 1 randomized pair of questions.
5. As a practical note, one should instruct the respondent to write down his or her answer to the sensitive question on a slip of paper (to be discarded afterward) before rolling the die. Otherwise, any delay in responding may suggest to the interviewer that the respondent's roll directed him or her to answer the threatening question.
6. As a further extension, Su, Sedory, and Singh (2014) proposed a two-stage version of the Kuk design. Rather than responding based on a single-stage draw of cards, the outcome of that draw would direct the respondent to manipulate one of two additional randomizing devices (with differing probability values) to determine the appropriate True/False answer. The Su et al. design may increase efficiency without compromising privacy, but the additional step in the process may try the patience of some respondents.