Key Results About Expectations and Variances of Random Variables With Applications for Sampling From Finite Populations (Revision)

Stephen E. Fienberg
Department of Statistics
Carnegie Mellon University

March 11, 2003

There are a number of basic formulas for the determination of expectations and variances of random variables that we will use repeatedly in this course. Virtually all students in the class have been exposed to then before in other courses, but you may not recall all of the details. The purpose of this handout is to review these results in a single convenient place for your reference. We’ll also consider a few illustrative examples applying the ideas.

1 Definitions of Expectations and Variances

Let $X$ be a discrete random variable taking values $x_i$ for $i = 1, 2, \ldots, k$ with probabilities $p_i$. Then

$$\sum_{i=1}^{k} p_i = 1,$$

$$\mu_X = E[X] = \sum_{i=1}^{k} x_i Pr(X = x_i) = \sum_{i=1}^{k} x_i p_i,$$

$$\sigma_X^2 = Var[X] = \sum_{i=1}^{k} (x_i - E[X])^2 Pr(X = x_i) = \sum_{i=1}^{k} (x_i - \mu_X)^2 p_i.$$

More generally, as we saw in Appendix B of Lohr (1999),

$$E[g(X)] = \sum_{x} g(x) P(X = x),$$

and the results for linear functions below follow from this more general definition.

[The same formulas hold for continuous or discrete random variables and we replace the summation in formulae (1) through (3) by integral signs. For sampling from finite populations for either attributes or measurements we need not deal with this complication except for our use implicitly of normal approximations.]
The following formulas, which we have been using in class and in homework assignments, are a direct application of these definitions. Let \( \overline{y} \) be the average of a sample of size \( n \) taking \( k \) different values \( \overline{y}_i \) for \( i = 1, 2, \ldots, k \), and let \( \overline{y}_U \) be the mean of all of the elements in the finite population of size \( N \). Then

\[
E[\overline{y}] = \sum_{i=1}^{k} \overline{y}_i \Pr(\overline{y} = \overline{y}_i),
\]

\[
Var[\overline{y}] = \sum_{i=1}^{k} (\overline{y}_i - E[\overline{y}])^2 \Pr(\overline{y} = \overline{y}_i),
\]

\[
Bias[\overline{y}] = E[\overline{y}] - \overline{y}_U,
\]

\[
MSE[\overline{y}] = Var[\overline{y}] + (Bias[\overline{y}])^2.
\]

We repeat these formulae here for convenience and in the following sections we will provide tools for their calculation in various special cases of interest such as for simple random sampling (SRS) with and without replacement (i.e., finite populations and infinite populations), both of which we discuss in detail.

In a separate handout, to be distributed later, we will deal with the case of stratified random sampling.

2 Expectations and Variance for Multiples and Sums

2.1 Working With One Random Variable

Suppose we have a random variable \( X \) with expectation \( \mu \) and variance \( \sigma^2 \), and let \( a \) and \( b \) be any constants (e.g., such as \( 1 \)). Then it is straightforward to show that:

\[
E[X + b] = E[X] + b = \mu + b,
\]

\[
Var[X + b] = Var[X] = \sigma^2,
\]

\[
E[aX] = aE[X] = a\mu,
\]

\[
Var[aX] = a^2Var[X] = a^2\sigma^2.
\]

Thus we can show that

\[
E[aX + b] = aE[X] + b = a\mu + b.
\]

\[
Var[aX + b] = a^2Var[X] = a^2\sigma^2.
\]

2.2 Example 2: Expectation and Variance for SRS and Attributes

Suppose we are taking a simple random sample (SRS) of size \( n \) from a population of size \( N \) without replacement and are interested in the proportion \( p = N_A/N \) of units with attribute A where \( N_A \) is the number of units with attribute A.

We know from class (and Appendix B of Lohr, 1999) that the distribution of \( X \), the number of sample units with attribute A, is hypergeometric:

\[
Pr(X = x|n, N_A, N) = \frac{\binom{N_A}{x} \binom{N-N_A}{n-x}}{\binom{N}{n}}.
\]
Exercise: Calculate $E[X]$ and $E[X(X-1)]$ and hence show that the sample average, $\bar{p} = X/n$, is unbiased and has variance

$$\text{Var}[\bar{p}] = \frac{p(1-p)}{n} \left( \frac{N-n}{N-1} \right).$$

(16)

2.3 Example 2: Randomized Response

2.3.1 What is Randomized Response? (From Lohr, 1999, p. 404)

Sometimes you want to conduct a survey asking very sensitive questions such as “Do you use Cocaine?” or “Have you ever shoplifted?” or “Did you understate your income on your tax return?”

These are all questions that “yes” respondents could be expected to lie about. A question form that encourages truthful answers but makes people comfortable is desired. Horvitz, et al. (1967), in a variation of Warner’s (1965) original data, suggest using two questions—the sensitive question and an innocuous question—and using a randomizing device (such as a coin flip) to determine which question the respondent should answer. If a coin flip is used as the randomizing device, the respondent might be instructed to answer the question “Did you use Cocaine in the past week?” if the coin is heads, and “Is the second hand on your watch between 0 and 30?” if the coin is tails. The interviewer does not know whether the coin was head or tails and hence does not know which question is being answered. It is hoped that the randomization and the knowledge that the interviewer does not know which question is being answered will encourage respondents to tell the truth if they have used cocaine in the past week.

The randomizing device can be anything, but it must have a known probability $P$ that the person is asked the sensitive question and probability $1-P$ that the person is asked the innocuous question. We’ll discuss other forms of the randomized response model in class.

2.3.2 Estimating $\pi$, the Proportion of “yes” Responses to the Sensitive Question

The key to randomized response is that the probability that the person responds “yes” to the innocuous question is known. We want to estimate $\pi$, the proportion responding “yes” to the sensitive question.

Denote by $\lambda$ the real probability of replying “yes”. This probability can be divided into two parts, using the law of total probability:

$$\lambda = Pr(Yes) = Pr(Yes \mid \text{Sensitive Question}) \times Pr(Answer \text{ sensitive question}) +$$

$$+ Pr(Yes \mid \text{Innocuous Question}) \times Pr(Answer \text{ Innocuous Question})$$

$$= \pi \times P + Pr(Yes \mid \text{Innocuous Question}) \times (1-P)$$

(17)

Since we usually have an estimate of $\lambda$, which is the proportion of “yes” answers in the sample, we can use it to estimate $\pi$:

$$\hat{\pi} = \frac{\lambda - Pr(Yes \mid \text{Innocuous Question}) \times (1-P)}{P}.$$  

(18)

2.3.3 An Example

We want to estimate the proportion of students that drink alcoholic beverages on CMU campus ($\pi$). For that purpose each of the students in class is presented with two questions:
• Question A: “Have you had an alcoholic beverage on the Carnegie Mellon campus during the last two weeks?”

• Question B: “Is the last digit of your Social Security number odd?” (1,3,5,7,9)

The randomizing mechanism is an urn with 80 blue balls and 20 yellow balls. Draw a ball at random. If it is blue, answer Question A. If it is yellow, answer Question B.

In this example,

\[ P = 0.8 \]

\[ Pr(Yes \mid \text{Innocuous Question}) = 0.5 \, . \]

From equation (19) we get:

\[ \hat{\pi} = (\hat{\lambda} - 0.1)/0.8 \, . \]  \hfill (19)

2.3.4 What are the Mean and Variance of \( \hat{\pi} \)?

Remember that in an SRS, \( \hat{\lambda} \), which is the proportion of “yes” answers in the sample, is an unbiased estimator of \( \lambda \) (the proportion of “yes” answers in the population). \( P \) and \( Pr(Yes \mid \text{Innocuous Question}) \) are constants. Hence, using rule (13) for expectations we get:

\[ E(\hat{\pi}) = \frac{1}{P} \left( E(\hat{\lambda}) - Pr(Yes \mid \text{Innocuous Question}) \right) = (\lambda - 0.1)/0.8 = \pi. \]  \hfill (20)

This means that \( \hat{\pi} \) is an unbiased estimator for estimating \( \pi \). What about the variance?

Using rule (14) for the variance, we get:

\[ V(\hat{\pi}) = \frac{1}{P^2} Var(\hat{\lambda}), \]  \hfill (21)

and we know that in an SRS:

\[ V(\hat{\lambda}) = \frac{\lambda(1-\lambda)}{n} \frac{N-n}{N-1}. \]  \hfill (22)

Since \( \lambda \) is unknown, we can obtain an estimator of \( Var(\hat{\pi}) \), by plugging \( \hat{\lambda} \) instead of \( \lambda \) in the previous equation. Finally, we get:

\[ V(\hat{\pi}) = \frac{1}{.64} \frac{\hat{\lambda}(1-\hat{\lambda})}{n} \frac{N-n}{N-1}. \]  \hfill (23)

2.4 Two Random Variables

Now we consider two random variables \( X_1 \) and \( X_2 \), and let \( a_1 \), and \( a_2 \) be any constants (e.g., such as 1). Let \( X_i \) have expectation \( \mu_i \) and variance \( \sigma_i^2 \) for \( i = 1,2 \). Then if we recall that the expectation of the sum of two random variables is the sum of their expectations, i.e.,

\[ E[X_1 + X_2] = E[X_1] + E[X_2] = \mu_1 + \mu_2, \]  \hfill (24)

we can use the earlier results to show that

\[ E[a_1 X_1 + a_2 X_2] = a_1 E[X_1] + a_2 E[X_2] = a_1 \mu_1 + a_2 \mu_2. \]  \hfill (25)

Finally if \( X_1 \) and \( X_2 \) are independent,

\[ Var[X_1 + X_2] = Var[X_1] + Var[X_2] = \sigma_1^2 + \sigma_2^2, \]  \hfill (26)
and hence
\[
\text{Var}[a_1X_1 + a_2X_2] = a_1^2\text{Var}[X_1] + a_2^2\text{Var}[X_2] = a_1^2\sigma_1^2 + a_2^2\sigma_2^2. \tag{27}
\]

Note that, by setting \(a_1 = 1\) and \(a_2 = -1\), this result implies
\[
\text{Var}[X_1 - X_2] = \text{Var}[X_1] + \text{Var}[X_2] = \sigma_1^2 + \sigma_2^2. \tag{28}
\]

Unfortunately, when we sample from a finite population without replacement, the successive draws from the population have changing probabilities and thus the elements of a sample and not independent of one another. To evaluate the variance of a sum when two random variables are not independent we need to introduce the covariance.

Let \(X_1\) take the \(k_1\) values \(x_{1i}\) and \(X_2\) take the \(k_2\) values \(x_{2j}\). then the covariance is:

\[
\text{Cov}[X_1, X_2] = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} x_{1i}x_{2j}P(r(X_1 = x_{1i} \text{ and } X_2 = x_{2j}) - \mu_1\mu_2). \tag{29}
\]

and
\[
\text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] + 2\text{Cov}[X_1, X_2], \tag{30}
\]

and hence
\[
\text{Var}[a_1X_1 + a_2X_2] = a_1^2\text{Var}[X_1] + a_2^2\text{Var}[X_2] + 2a_1a_2\text{Cov}[X_1, X_2]. \tag{31}
\]

When the two random variables, \(X_1\) and \(X_2\), are in fact independent \(\mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]\) and \(\text{Cov}[X_1, X_2] = 0\) and these formulas reduce to those above.

### 2.5 Example 3: Repeated Polls

The CNN poll taken on June 1996 of 818 registered voters showed Clinton would get 54% of the votes. A similar pole that was taken a week earlier had 57% voters for Clinton. We are interested in estimating the difference, to see whether there was a change in the entire population percent of Clinton voters. Let us denote by \(\hat{p}_2\) the proportion of Clinton voters on the early poll, and \(\hat{p}_1\) on the later poll. Then, a natural estimator of the real difference \(p_2 - p_1\) is \(\hat{p}_2 - \hat{p}_1 = 3\%\).

Is this estimator biased? We know that \(\hat{p}_1\) is unbiased for estimating \(p_1\), and \(\hat{p}_2\) is unbiased for estimating \(p_2\). From (25) we get
\[
\mathbb{E}(\hat{p}_2 - \hat{p}_1) = \mathbb{E}(\hat{p}_2) - \mathbb{E}(\hat{p}_1) = p_2 - p_1 \tag{32}
\]

We have an unbiased estimator for the difference.

What is the variance of this estimator? Assuming that the polls are independent, we can use (28):
\[
\text{Var}(\hat{p}_2 - \hat{p}_1) = \text{Var}(\hat{p}_2) + \text{Var}(\hat{p}_2) = \frac{p_2(1 - p_2)}{818} + \frac{p_1(1 - p_1)}{818} \tag{33}
\]

(why did we drop out the finite population correction for the variances?)

We can only get an estimate of this variance, by plugging \(\hat{p}_1, \hat{p}_2\) instead of \(p_1, p_2\) in the previous equation. We then get:
\[
\text{Var}(\hat{p}_2 - \hat{p}_1) = \frac{.54 \times .46}{818} + \frac{.57 \times .43}{818} = .006 \tag{34}
\]

We can now construct a 95% confidence interval for the difference in Clinton votes:
\[
3\% \pm 1.96 \times 0.06\% = [2.88\%, 3.11\%] \tag{35}
\]
hence we conclude that the difference is significant at a 5% significance level.
2.6 \( n \) Random Variables

The results for linear combinations of random variables follow direct from those for two in the preceding subsection. We can consider \( n \) random variables \( \{ X_i \} \), where \( X_i \) has mean \( \mu_i \) and variance \( \sigma_i^2 \). Further let \( a_i \) for \( i = 1, 2, \ldots, n \) be constants. Then the expectation of a linear combination of random variables is the linear combination of the expectations:

\[
E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i]. \tag{36}
\]

The variance of a linear combination of random variables in general involves both their variances and covariances:

\[
Var[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i^2 Var[X_i] + 2 \sum_{i>j} a_i a_j Cov[X_i, X_j]. \tag{37}
\]

But as in the case of \( n = 2 \), if the random variables are independent the covariance are all zero and we get the simpler result that:

\[
Var[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i^2 Var[X_i]. \tag{38}
\]

3 Formulas for the Expectation and Variance of an Average

We are now in a position to really put our formulas for expectations and variances to work in the finite population setting. Suppose we have a finite population of size \( N \) of measurements and with population mean

\[
\mu = \frac{\sum_{i=1}^{N} X_i}{N} \tag{39}
\]

and population variance

\[
\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N} \tag{40}
\]

We will work here in this section only with measurements but attributes (whose results we derived above as an exercise) are really just a special case where each measurement takes the values 1 or 0.

Let’s begin with expectations. If the we take a simple random sample of size \( n \), whose corresponding random variables are \( \{ X_i \} \). The values for each member of the sample are unknown in advance and the finite population makes them dependent. But before we see any of their values they each have the same expectation \( \mu \), i.e., the population mean. Therefore we can evaluate the expectation of the the sample mean, \( \bar{X} = \sum_{i=1}^{n} X_i / n \), using formula (36) with \( a_i = 1/n \):

\[
E[\bar{X}] = E \left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu. \tag{41}
\]

This is just the property of unbiasedness of the sample mean which we discussed earlier in class. If we let \( N \) get large then we get sampling with replacement as the limiting case.
The result for the variance of the sample mean from a finite population uses formula (37) and the fact that the variances of \( \{X_i\} \) are all the same as are the covariances representing the dependence induced by the finite population size for all 
\[
\binom{N}{2} = \frac{N \times (N - 1)}{2}
\]
possible pairs of random variables making up the sample. Thus
\[
\text{Var}[\bar{X}] = \text{Var}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]
\]
\[
= \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}[X_i] + \frac{2}{n^2} \sum_{i>j} \text{Cov}[X_i, X_j]
\]
\[
= \frac{\sigma^2}{n} + \frac{n-1}{n} \rho \sigma^2
\]
\[
= \frac{\sigma^2}{n} [1 + (n - 1) \rho],
\]
where \( \rho \) is the correlation between a pair of sample items. It turns out the \( \rho \) depends only on \( N \), not on \( n \), since we are essentially asking about the correlation between a pair of units drawn from a finite population. In fact,
\[
\rho = -\frac{1}{N - 1}
\]
and so equation (45) reduces to our formula for the variance of a sample mean for SRS without replacement:
\[
\text{Var}[\bar{X}] = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1}\right).
\]

The way that statisticians usually deal with most calculations involving sampling from a finite population is by introducing special random variables for the selection of units. For \( i = 1, 2, \ldots, N \), we let \( Z_i \) be 1 if the \( i \)th unit is in the sample and 0 if it is not. Then, as Lohr (1999) shows in Appendix B, \( P(Z_i = 1) = \frac{n}{N} \) and hence
\[
E(Z_i) = E(Z_i^2) = 0 \times P(Z_i = 0) + 1 \times P(Z_i = 1) = \frac{n}{N},
\]
and for \( i \neq j \),
\[
E(Z_i Z_j) = \frac{n(n-1)}{N(N-1)}.
\]
We can thus show that
\[
\text{Var}(Z_i) = \frac{n(N-n)}{N^2}
\]
and for \( i \neq j \),
\[
\text{Cov}(Z_i, Z_j) = -\frac{n(N-n)}{N^2(N-1)}.
\]
Hence we get that \( \rho = -\frac{1}{N-1} \).
Finally, we note that when the finite population size $N$ gets large, the correlation term goes to zero, and the last term in equation (46) tends to 1. Thus the we get the usual formula for the variance of a sample mean for SRS with replacement

$$Var[\bar{X}] = \frac{\sigma^2}{n},$$

which also follows directly from equation (38).

4 Some Other Uses for Our Formulas

Stratified random sampling. In a separate handout we will use equations (36), (37), and (38) to determine the means and variances of stratified random samples. There our population of $N$ units is divided into subpopulations or strata of sizes $N_i$, where $\sum_i N_i = N$, and we likewise divide our sample of size $n$ into separate pieces, $n_i$ where $\sum_i n_i = n$, one for each stratum. Thus we take an SRS without replacement from each stratum, and these samples from the separate strata are independent.

Comparisons between populations or subpopulations. Next suppose we take two sub-populations or strata and want to make inferences about the difference of the subpopulation means. We do this by looking at the difference between the sample means from these strata. This involves a different but again direct application of equations (36), (37), and (38), as well as equation (28) for the variance of the difference of two random variables.

Note: The stratified sampling handout will talk about “optimal choices” of the strata sample sizes, \( \{n_i\} \), to make the variance of the estimate of the overall population as precise as possible. Many groups will need to think about the best way to choose sample sizes for a stratified random sample in order to get the most precise comparisons of means.

References

