Simulation Analyses of the Horvitz-Thompson Statistic for Estimating Total Fatalities in the Altamont

By Julie Yee
September 10, 2009

Purpose
Recent reports and publications which estimate total fatalities in the Altamont have utilized a method which applies an expansion factor on the number of fatalities that have been counted. This method which originated as the Horvitz-Thompson estimator in sampling theory is now widely used by research biometricians in the wildlife sciences (Horvitz and Thompson, 1952; Cochran, 1977; Steinhorst and Samuel, 1989; Williams et al, 2002). Although wildlife biometricians generally use this method for estimating the size of live animal populations, it is easy to see its applicability to the problem of estimating total fatalities in the Altamont. However, the complex process of monitoring and counting fatalities at the Altamont has raised concerns about the efficacy of this estimator. This is in large part because the sighting probability of a carcass, which is a critical component of the estimator, is subject to uncertain errors. Simulation analyses can be used to demonstrate the performance of an estimation method under computer-controlled conditions in which the total fatalities is known. Briefly, the computer generates count data by randomly simulating the process of carcass removal and observer detection through probability selection. By applying the Horvitz-Thompson estimator to the simulated count data and comparing to the known total fatalities, then these simulation analyses allows us to assess the performance of that estimator (i.e. in terms of bias and precision). This document demonstrates simulation analyses for the trivial situation where the sighting probability (based on carcass removal and observer detection processes) is precisely known. I plan to perform future simulations under more liberal and realistic conditions in which the sighting probability is unknown and subject to error.

Estimator
Williams et al (2002, page 256) presents a general form of the Horvitz-Thompson estimator as

$$\hat{N} = \sum_{i=1}^{C} \frac{1}{\beta_i} \qquad (1)$$

where the hat symbol (^) distinguishes the estimated total fatalities (i.e., $\hat{N}$) from the actual total fatalities (i.e., N), C is the number of counted fatalities, and β_i is the sighting probability for the i^{th} fatality count. Note that if the sighting probability is equal for all fatalities, then the estimator simplifies to

$$\hat{N} = \frac{C}{\beta}. \qquad (2)$$

The majority of documents reviewed by the SRC which estimate Altamont fatalities have used either equation (1) or (2). Most recently the Monitoring Team has applied (1) to Altamont fatalities using a sighting probability of $\hat{\beta}_i = \hat{R}_{Ci} \times \hat{p}_i$, where $\hat{R}_{Ci}$ is the estimated probability of the carcass remaining in the search area after being deposited anytime since the previous search where the carcass was found and $\hat{p}_i$ is the probability the carcass is detected by an observer given that it is still present at the time of the search. This is the same estimator I used for the simulations.

<u>Simulations</u>

This exercise simulates the estimation of fatalities at 2500 surveyed turbines. Although in practice, Altamont-wide estimates requires extrapolating to 5000 or some-odd total turbines, I have only performed this simulation to estimate total fatalities at turbines where surveys occur. I simulated data at the turbine-day scale for one year, over 2500 turbines and 365 days, or 912,500 simulation replicates. For each replicate, I generated a random number of total fatalities [following a Poisson random number generator with mean=0.0001 which equates to a rate of 91.25 total carcasses per year per 2500 turbines]. The majority of turbines had zero fatalities. To simulate some of the irregularities in the monitoring process, search dates were selected at randomly varying intervals between 25-50 days apart [following a uniform (25-50) random number generator]. Within these intervals, carcasses were allowed to disappear before the next search date occurred. Carcasses disappeared according to the probability $\hat{R}_{Ci}$ which I calculated from the estimate $\hat{R}_i = 1 - 0.2062888 \times \log(day)$ (document M-32), following Smallwood (2007) (document R-50) definitions for $\hat{R}_{Ci}$ and $\hat{R}_i$. For simplicity, I used a detection probability $\hat{p}_i$=1, so that any carcass that remained is assumed to have been counted (note: I will generalize this to estimates of unknown *p* in future simulations). Thus, in this manner, I simulated a set of count data with which I could assess the performance of the Horwitz-Thompson estimator. I repeated this simulation for 1000 iterations to examine, in general, how the estimator performs.

<u>Results</u>

Results of the simulation on pages 4-5 show the following:

adjustedx = adjusted fatality count, same as $\hat{N}$ Horvitz Thompson estimator, shown as an average across iterations and by individual iteration

error = $\hat{N}_i - N_i$

fallen = total fatalities (fallen birds), same as *N*, shown as average and by iteration

found = counted fatalities (found birds), shown as an average and by iteration

iter = iteration number

num_iterations = 1000 simulations iterated

squared error = $\left(\hat{N}_i - N_i\right)^2$

mean_squared_error = $\frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{N}_i - N_i\right)^2$, for use as an estimate of error variance

root_mean_squared_error = $\sqrt{\frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{N}_i - N_i\right)^2}$, for use as estimated standard error

Page 4 contains tabulated results. On average, 92.276 fatalities were simulated. By the "scavenger removal" process, a mean of 40.646 fatalities were counted. The Horvitz-Thompson estimator averaged 92.460 fatalities. The bias (92.460 – 92.276 = 0.184) is virtually negligible suggesting that the estimator is on average accurate. The standard error is 10.9582 or slightly more than 10% of the number of fatalities, which suggests that most times the Horvitz-Thompson estimator in this setting can overestimate or underestimate total fatalities by up to 20%.

References

Cochran, W.G. 1977. Sampling Techniques, Third Edition. John Wiley & Sons, Inc., New York, NY. 428 pp.

Horvitz, D.G. and D.J. Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association* 47:663-685.

Smallwood, K.S. 2007. Estimating wind turbine-caused mortality. *Journal of Wildlife Management*. 2781-2791.

Steinhorst, R.K. and M.D. Samuel. 1989. Sightability adjustment methods for aerial surveys of wildlife populations. *Biometrics* 45:415-425.

Williams, B.K., J.D. Nichols, M.J. Conroy. 2002. *Analysis and Management of Animal Populations: Modeling, Estimation, and Decision Making*. Academic Press. San Diego, CA. USA.

Summary results:

num_iterations	fallen	found	adjustedx	bias	mean_squared_error	root_mean_squared_error
1000	92.276	40.646	92.459805325	0.18381	120.083	10.9582

Results by iteration (first 40 only):

iter	fallen	found	adjustedx	error	squared_error
1	96	51	116.45667936	20.4567	418.48
2	104	42	98.462276779	-5.5377	30.67
3	94	34	79.096207103	-14.9038	222.12
4	93	32	72.700092041	-20.2999	412.09
5	103	44	98.911360143	-4.0886	16.72
6	94	45	103.32592366	9.3259	86.97
7	79	37	82.752790815	3.7528	14.08
8	97	40	89.045872282	-7.9541	63.27
9	92	36	82.373038926	-9.6270	92.68
10	82	31	71.707902974	-10.2921	105.93
11	101	39	90.708372831	-10.2916	105.92
12	97	36	82.399289275	-14.6007	213.18
13	91	43	98.252644073	7.2526	52.60
14	89	39	89.338913061	0.3389	0.11
15	97	38	88.6188541	-8.3811	70.24
16	100	42	94.544287135	-5.4557	29.76
17	110	45	102.78450625	-7.2155	52.06
18	94	37	83.600492802	-10.3995	108.15
19	92	37	83.676517746	-8.3235	69.28
20	101	42	92.855380862	-8.1446	66.33
21	93	41	93.933268643	0.9333	0.87
22	116	48	107.75479198	-8.2452	67.98
23	99	41	93.409921418	-5.5901	31.25
24	105	46	105.52701771	0.5270	0.28
25	83	36	81.930860372	-1.0691	1.14
26	102	43	97.433046324	-4.5670	20.86
27	93	38	86.717219962	-6.2828	39.47
28	84	42	95.024970217	11.0250	121.55
29	93	44	99.018754127	6.0188	36.23
30	92	31	70.7643029	-21.2357	450.95
31	89	41	93.194837304	4.1948	17.60
32	84	33	75.623884615	-8.3761	70.16
33	88	32	72.734591423	-15.2654	233.03
34	108	43	97.167431799	-10.8326	117.34
35	98	39	85.703003099	-12.2970	151.22
36	78	37	84.135751507	6.1358	37.65
37	78	34	76.009623271	-1.9904	3.96
38	101	45	101.82908155	0.8291	0.69
39	85	45	102.37481842	17.3748	301.88
40	93	33	74.567453242	-18.4325	339.76