INVESTIGATION OF EMPIRICAL BAYESIAN CONFIDENCE INTERVALS FOR EXPENDITURE DATA IN THE MEDICAL EXPENDITURE PANEL SURVEY

Robert Baskin[1], Malay Ghosh[2]
[1]AHRQ
[2]University of Florida

1. Introduction:

If an end user of survey data from a complex sample is interested in calculating a confidence interval for a mean, total, or proportion, then typically the justification of approximately valid confidence intervals is based on the central limit theorem. Validity of this normal approximation for a given sample size is typically determined by rules of thumb usually based on Edgeworth expansions. Cochran (1977) gives a rule of thumb for determining sample size if the principle problem in the data is skewness. Cochran's estimate of the sample size necessary to achieve nominal coverage is based on the Fisher's skewness coefficient, G_1. Cochran's rather simple rule is the sample size should be at least $25G_1^2$ in order for a 95% confidence interval to have a coverage rate of at least 94%.

Most nationally representative surveys have sufficient sample size by Cochran's criteria to produce valid confidence intervals for national or even regional level estimates of means or totals. However, analysts are often interested in subdomain specific estimates and even for large national surveys, subdomains of interest can be sparse. Since the finite population intervals are based on normal approximations to the sampling distribution, and if the sample size cannot be changed, the only alternatives are to either rely on the distributional assumptions or to use a confidence interval correction. Of course, if an analyst is willing to build models, then quite accurate estimates might be constructed. But if an analyst simply wants to construct a confidence interval for skewed data within a sparse subdomain the question is whether distributional assumptions can increase the accuracy of confidence interval coverage. This paper will attempt to address this issue in the setting of stratified sampling.

Generally, attempts to improve the accuracy of confidence interval coverage have proceeded along the lines of Edgeworth corrections, saddlepoint approximations, bootstrap corrections, corrections to Empirical Bayes intervals, or generalized confidence intervals. This paper investigates the Empirical Bayes intervals and compares them to the standard normal intervals and ad hoc gamma intervals.

This paper will proceed as follows. Section two addresses constructing confidence intervals based on the gamma distribution. In section three, we discuss intervals based on the lognormal distribution. The lognormal interval attributed to Cox will be presented. Section four describes the Empirical Bayes interval based on a lognormal distribution. Section five presents the results of two types of simulations. The first type of simulation is based on simple random sampling from a pseudo-population of skewed data. The second type of simulation is based on stratified sampling from the pseudo-population of skewed data. Results are given in section six. Finally, the last section provides a conclusion with directions for future research.

2. Gamma Confidence Intervals

2.1 The Gamma Distribution for a Sample Mean

The process of constructing gamma intervals is actually straightforward given that a gamma distribution has a multiplicative location parameter. In this paper, if X has a gamma distribution with shape parameter, ν, and scale parameter, λ, it will mean that X has density proportional to $x^{\nu-1}e^{-x/\lambda}$ on the non-negative real numbers. This is denoted as X ~ $\Gamma(\nu,\lambda)$ and under this parameterization $E(X)=\nu\lambda$ and $Var(X)=\nu\lambda^2$. A simple calculation using the moment generating function shows that for a simple random sample of size n from a gamma distribution, and constructing the sample mean as $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$, then $\overline{X}$ ~ $\Gamma(n\nu,\lambda/n)$.

2.2 Construction of Simple Gamma Confidence Intervals

For notation, assume that α is the nominal confidence level of interest and μ is the population mean which we are trying to estimate. Regardless of what the underlying distribution is, $E(\overline{X}) = \mu$, the population mean, and this is the basis of both normal and gamma confidence intervals. To construct a gamma confidence interval, start with the assumption that $\overline{X}/\mu$ is distributed as gamma with mean 1. Note that if $\overline{X}/\mu$ is distributed as gamma with mean 1, then since the mean of a gamma distribution is the product of the scale parameter times the shape parameter it must be that $\overline{X}/\mu$ is distributed as $\Gamma(n\nu, 1/n\nu)$, i.e., $\lambda = 1/n\nu$. Under the assumption that $\overline{X}/\mu$ has a gamma distribution, and assuming that ν is known, the probability statement $P\{\Gamma(\alpha/2; n\nu, 1/n\nu) \le \overline{X}/\mu \le \Gamma(1-\alpha/2; n\nu, 1/n\nu)\} = 1-\alpha$ is rearranged to produce the confidence interval $P\{\overline{X}/\Gamma(1-\alpha/2; n\nu, 1/n\nu) \le \mu \le \overline{X}/\Gamma(\alpha/2; n\nu, 1/n\nu)\} = 1-\alpha$. Since ν is rarely known it must be estimated. If an estimate, $\hat{\nu}$, is used in the formula then the distribution is only approximately gamma and the coverage level is only approximate. The small sample exact distribution of the gamma interval with an estimated shape is unknown but the simulation results indicated satisfactory coverage with using estimated shape.

The above interval is one of many $1-\alpha$ confidence intervals for the population mean under the assumption that the underlying distribution is gamma. However, it is not "best" in the sense of shortest $1-\alpha$ interval, but for the current examination this simple interval is sufficient. A small simulation indicated that under simple random sampling there was minimal difference in the performance of this equal tail coverage gamma interval and the shortest width gamma interval.

The derivation of the gamma distribution for the sample mean can also be done under more general assumptions of a stratified sample. If we assume that the strata means have a gamma distribution with common shape parameter, ν, strata unique scale parameters, $\lambda(i)$, and that the strata are independently sampled, then the overall sample mean is distributed as gamma using a similar derivation as for the simple random sample. The shape parameter of the overall mean will be equal to the common shape parameter from the strata divided by the number of strata. The scale parameter of the overall mean will be equal to the sum of the strata level scale parameters. Of course, this would be a weighted sum if the overall mean were a weighted sum of the strata means.

The assumption that the shape parameters in the strata are approximately equal is equivalent to the idea each strata has approximately the same coefficient of variation. From a survey samplers point of view this might be a workable assumption if the survey were designed specifically to collect the skewed variable and the strata were designed and allocated relative to the variance of the skewed variable of interest. Unfortunately, this is generally not the case. Assumptions about a common shape parameter and estimating this shape parameter appear to be the main difficulty with using a gamma distribution. However, in simulations the gamma intervals using an estimated shape parameter seemed to perform satisfactorily and certainly better than normal intervals.

2.3 Estimation of the shape Parameter

The situation in which the shape parameter, ν, is known corresponds to the situation for a normal distribution in which σ is known but this rarely occurs in practice. Thus, ν must be estimated from some source, usually the sample, just like σ must nearly always be estimated from the sample. Using the relationships between the gamma parameters and the sample mean and sample standard deviation one can derive a simple method of moments estimator of ν as $\hat{\nu} = \overline{X}^2 / sd(X)^2$. This gives an interpretation of $\lambda = 1/\nu$ as a direct function of the coefficient of variation.

The gamma distribution with an estimated shape parameter can be thought of analogously to the Student t distribution for the normal distribution with an estimated standard deviation. The small sample distribution of a gamma with an estimated shape parameter is currently unknown and is an area for possible future research.

In this paper, the following idea is used based on the assumption that the total sample is relatively large but the interest is in writing confidence intervals for somewhat sparse subdomains. If there is truly a common shape parameter in each stratum, then that common shape parameter should also hold for the entire sample and the entire sample can be used to estimate the common shape parameter. The estimate of the shape from the entire sample can then be used in each subdomain of interest. This is not an ideal solution but seems to perform well in simulations. Note that the method of estimation for the shape parameter used here is a simple method of moments estimator but more sophisticated estimation techniques could be investigated. There are works on maximum likelihood estimators for estimating gamma parameters and robust parameter estimation for a gamma distribution. However, the ultimate goal is to provide the end user with a simple formula for a gamma confidence interval, so the simple method of moments estimation of the shape parameter is satisfactory for this purpose.

2.4 Summary on Gamma Intervals

This section provided an introduction to gamma based confidence intervals used in the simulations. Assuming the sample mean follows a gamma distribution has a theoretical basis for either simple random sampling or for stratified sampling. The simple gamma interval requires an estimate of the shape parameter which is the weakness of the approach. A small simulation indicated that the simple gamma interval and the normal interval were numerically very similar for subsample sizes of 5,000 or larger.

3. Lognormal Confidence Intervals

3.1 The Lognormal Distribution

The lognormal distribution can also be used as a distributional model for heavy tailed data. For example Zhou and Gao (1997) examine lognormal confidence intervals for health expenditure data. The work of Zhou and Gao indicates that lognormal confidence intervals work well for health expenditure data associated with knee replacement in a sample of size 355.

If X has a lognormal distribution with location parameter, μ, and scale parameter, σ, it means that log(X) is N(μ,σ) and that X has density proportional to $x^{-1}e^{-\frac{1}{2\sigma^2}(\log(x)-\mu)^2}$ on the positive real numbers. This will be denoted as X ~ $\Lambda(\mu,\sigma)$ and under this parameterization $E(X)=e^{\mu+\frac{1}{2}\sigma^2}$ and $Var(X)=e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$. The form of the expected value is suggestive that the parameter $\mu+\frac{1}{2}\sigma^2$ can be used to construct confidence intervals for lognormal data. In fact all of the lognormal intervals studied are based on first transforming the data by taking a log, second estimating $\mu+\frac{1}{2}\sigma^2$, third writing a confidence interval for $\mu+\frac{1}{2}\sigma^2$ based on some distributional assumption, and finally transforming the interval back to the original scale using the exponential function.

Note that if the individual observations are thought to be lognormal then the mean of the observations is not lognormally distributed. It is not clear if there is a theoretical basis for using a lognormal interval for the overall mean with stratified sampling. However, in the simulations it is included.

3.2 Construction of Lognormal Confidence Intervals

The interval proposed by Cox as reported in Land (1972) is the simplest interval of this form. If the original variable is denoted by X, then let Y denote the log of X. First, the Cox interval estimates $\mu+\frac{1}{2}\sigma^2$ by

$$\bar{Y}+\tfrac{1}{2}S_Y{}^2=\frac{1}{n}\sum_1^n Y_i+\tfrac{1}{2}\frac{1}{n-1}\sum_1^n (Y_i-\bar{Y})^2.$$

Second, since the variance of $\bar{Y}+\frac{1}{2}S_Y{}^2$ is $\frac{\sigma^2}{n}+\frac{1}{2}\frac{\sigma^4}{n-1}$ this is estimated by $\frac{S_Y{}^2}{n}+\frac{1}{2}\frac{S_Y{}^4}{n-1}$. Third, the distribution of $\bar{Y}+\frac{1}{2}S_Y{}^2$ is assumed to be normal so a (1-α)% confidence interval for $\mu+\frac{1}{2}\sigma^2$ is given by

$$[\bar{Y}+\tfrac{1}{2}S_Y{}^2-\Phi(1-\frac{\alpha}{2})\sqrt{\frac{S_Y{}^2}{n}+\tfrac{1}{2}\frac{S_Y{}^4}{n-1}};$$

$$\bar{Y}+\tfrac{1}{2}S_Y{}^2-\Phi(\frac{\alpha}{2})\sqrt{\frac{S_Y{}^2}{n}+\tfrac{1}{2}\frac{S_Y{}^4}{n-1}}]$$

where Φ denotes a standard normal distribution.

Finally this interval is back-transformed using the exponential function.

3.3 Summary on Lognormal Intervals

This section presented the Cox interval for the lognormal distribution. The interval is based on a logarithm transform of the data, constructing a normal confidence interval for the parameter, which equals the mean plus one half the variance of the transformed data, then back transforming the data using an exponential function.

4. Empirical Bayes Lognormal Confidence Intervals

4.1 The Empirical Bayes approach

The lognormal distribution can also be used as a distributional model for Empirical Bayes estimation. Again, if X has a lognormal distribution with location parameter, μ, and scale parameter, σ, then log(X) is N(μ,σ) and that X has density proportional to $x^{-1}e^{-\frac{1}{2\sigma^2}(\log(x)-\mu)^2}$ on the positive real numbers. Under this parameterization $E(X)=e^{\mu+\frac{1}{2}\sigma^2}$ and $Var(X)=e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$. The approach used here will be to construct an Empirical Bayes confidence interval for the parameter $\mu+\frac{1}{2}\sigma^2$ and transform the interval back to the original scale using the exponential function.

4.2 Estimation of Empirical Bayes Confidence Intervals

The fully Bayesian approach would specify a lognormal distribution for X given the parameters, (μ,σ), along with a prior distribution for μ, $\mu_i \sim N(\alpha,\tau^2)$ which depends on the hyperparameters (α,τ^2). For the moment assume that both (σ^2,τ^2) are known. Then the posterior distribution of μ is given by

$$p(\mu \mid x,\tau^2,\alpha)=\frac{f(x\mid\mu)g(\mu\mid\tau^2,\alpha)}{m(x\mid\tau^2,\alpha)}$$

where $m(x\mid\nu,\alpha)$ is the marginal distribution of the data given the hyperparameters. In the Empirical Bayes approach, the hyperparameters are estimated from the marginal distribution, $m(x\mid\nu,\alpha)$, usually using maximum likelihood estimators.

In particular, if the log of the observed strata means are assumed to be independent normally distributed, $\log(X_i)\sim N(\mu_i,\sigma^2/n_i)$ where n_i is the sample size, and $\mu_i \sim N(\alpha,\tau^2)$ then the marginal distribution of log(X), given the hyperparameter α, is $m(\log(X_i)\mid\alpha)\sim N(\alpha,\sigma^2/n_i+\tau^2)$ so the marginal maximum likelihood estimator of α is $\hat{\alpha}=\frac{1}{K}\sum_1^K \log(X_i)$. Then the estimated posterior for $\mu_i \mid \log(X_i),\hat{\alpha}$ is $N(B\hat{\alpha}+(1-B)\log(X_i),(1-B)\sigma^2)$ where $B=\frac{\sigma^2/n_i}{\sigma^2/n_i+\tau^2}$. Since the posterior for $\mu_i\mid\log(X_i),\hat{\alpha}$ is normal with known parameters, and still assuming that both (σ^2,τ^2) are known, a naïve 95% Empirical Bayes confidence interval for $\mu+\frac{1}{2}\sigma^2$ can be constructed. As discussed in section 3.5 of Carlin and Louis (2000), the naïve Empirical Bayes confidence interval will undercover but a method by Morris will correct the naïve interval. The next step is to estimate the parameters (σ^2,τ^2) from the data. σ^2 is estimated by the within strata variance while τ^2 is estimated as the between strata variance. Finally this interval is back-transformed using the exponential function.

4.3 Summary on Empirical Bayes Intervals

This section presented the intervals for the lognormal distribution using Empirical Bayes methodology. The interval requires maximum likelihood estimation of hyperparameters.

5. Simulation Methods

This section describes how the pseudo-population was constructed and how the sampling was done. A discussion of the results follows.

5.1 Construction of the Pseudo Population

Berk and Monheit (2001) have previously described the skewed nature of healthcare expenditure data among the civilian non-institutionalized population of the United States from the Medical Expenditure Panel Survey (MEPS). MEPS is collected to estimate healthcare expenditures and usage among the civilian non-institutionalized population. To obtain skewed data for this study, healthcare expenditure data from six MEPS full year person-level public use files, available from www.AHRQ.MEPS.gov were merged along with a stratification that spanned 1996 to 2001. The pseudo-population consisted of over 156,000 records obtained from MEPS full year consolidated files HC-012, HC-020, HC-028, HC-035, HC-050, and HC-060. These records were grouped into 125 strata. Strata and primary sampling unit information were obtained from the MEPS 1996-2002 pooled estimation linkage file HC-036. For each record the following variables were extracted: total annual healthcare expenditures; sample information such as strata, primary sampling unit, and weight; and demographic and geographic information such as age, race, sex and region.

All records in the MEPS Public Use Files with nonzero expenditure, whether imputed or not, were retained. Approximately 28,000 records in the time period 1996-2001 had zero expenditures for the year so after these records were dropped there remained over 128,000 records with nonzero health expenditures for the year. One problem of healthcare expenditures is dealing with the semi-continuous nature of the data. The zero expenditure category can create problems and for this study only the nonzero expenditures were used, but including the zero category is a possible topic for future research.

5.2 Construction of Subdomains

As mentioned in the introduction, the primary interest is in constructing confidence intervals for subdomains, but it is not straightforward how to simulate sampling in order to estimate confidence interval coverage for subdomains. There appear to be two choices, neither of which is ideal. Starting with a fixed size sample for the total sample from the pseudo-population, we can look at fixed subdomains within the total sample, e.g., black females in the South, but these subsets will be of random size. Alternatively, we can take a random subdomain, not based on demographics, but of a fixed subsample size out of the total sample. The random subdomain approach allows control of the subsample size of the subset but it doesn't reflect inherent differences that can be found in expenditures in subdomains determined by demographic characteristics. However, both of these approaches were investigated and the results are reported.

5.3 Method of Sampling

Given the pseudo-population of approximately 128,000 records with nonzero expenditures, two methods of sampling were used. In both methods, a total sample of approximately 26,000 records was selected. This sample size matches the number of nonzero records in the 2001 MEPS full year file and thus reflects the total number of records an analyst might have to work with in a typical full year data file. For given total sample size of 26,000, the first method of sampling was to take a simple random sample from the pseudo-population as has been done in previous studies. The simulations were run in R provided by the R Core Development Team (2005).

5.4 Simple Random Sampling

In the finite population setting, most of the research done thus far regarding the effect of skewness on confidence intervals has simulated simple random sampling.

The approach of simple random sampling is included in this work to provide a validation for the simulation methodology and for the Cochran rule of thumb. To start with, a large sample of approximately 26,000 records, representing a full year file was drawn without replacement from the pseudo-population of 128,000 records. Then random subsamples of different sizes were drawn from the large sample full year file representation. For each fixed subsample size, the process of randomly subsampling the large sample was repeated 10,000 times. For each repetition of the subsampling process, the different intervals of interest were constructed and a count incremented for each confidence interval that covered the true pseudo-population mean. In one case, a subsample size starting at 20 and increasing to 1,000 by 10 was drawn. In another case, subsamples starting at 50 and increasing by 50 to 2,000 were drawn from the large sample.

5.5 Stratified Sampling

The effect of stratified sampling on confidence interval coverage is studied using simulations similar to the simple random sampling case. The common stratification from the public use file HC-036 provides a stratification for all records in the pseudo-population. The overall sample size of approximately 26,000 was used again. To allocate this sample size to the strata, the strata sample sizes from the MEPS 2001 data were used because the overall sample size was determined by that file. Given the sample size allocated to each stratum, a simple random sample within each stratum was selected from the pseudo-population. Since most national surveys such as MEPS are cluster samples the next step for future work should be to investigate cluster sampling within strata. `

The approach of writing confidence intervals from stratified sampling is important to evaluate for national samples and this work uses simulation to investigate the impact of stratified sampling on confidence intervals using the same pseudo-population. To start with, a large sample of approximately 26,000 records, representing a full year file is drawn by stratified sampling, without replacement, using simple random sampling within strata, from the pseudo-population of 128,000 records. This is facilitated by the ‘pps’ function in R, see Gambino (2003). After the stratified sample was selected, two methods of evaluating subdomains were investigated.

First, random subsamples of different sizes were drawn from the large sample full year file representation. For each fixed subsample size the process of randomly subsampling the large sample was repeated 10,000 times. For each repetition of the subsampling process the different intervals of interest were constructed and a count incremented for each confidence interval that covered the true pseudo-population mean. In one case a subsample size starting at 20 and increasing to 1,000 by 10 was drawn. In another case subsamples starting at 50 and increasing by 50 to 2,000 were drawn from the large sample.

Secondly, for certain fixed subdomains, such as crosses of sex, race-ethnicity, and region a stratified sample of approximately 26,000 was repeatedly drawn from the pseudo-population. A simulation of 10,000 repetitions of stratified sampling is reported in the Results section. Confidence intervals from each sample were constructed for each subdomain determined by the crossed variables. Then a count was incremented for each subdomain interval that covered the true pseudo-population mean.

6. Results

In several repeats of the simple random sample simulation, the total sample of 26,000 had a Fisher G_1 value of 10 to 11 indicating, according to Cochran’s rule, a sample size of over 2,000 would be necessary for a normal confidence interval with nominal value 0.95 to cover 94% of the time. This agreed with the observed results from Yu and Machlin. It was observed that the underlying pseudo-population has a Fisher G_1 of 13.9, so the simple random samples are consistently underestimating the skewness coefficient of the pseudo-population.

Figure I, shows coverage rates for the normal interval, the Cox lognormal interval, and the simple gamma interval. The simple gamma interval had the shape parameter estimated from the total sample as opposed to the subsample. The dotted horizontal line is the 94% line and the normal coverage rate does not approach the line until the subsample size starts to reach 2,000, which appears to agree with Cochran’s rule of thumb. The gamma coverage rate is consistently above the 94% line. The lognormal based intervals are initially good and this agrees with previous results from Zhou and Gao, for example. However, as the subsample size approaches 4,000 the lognormal coverage rates become worse than the normal and as the sample size increases the lognormal intervals get consistently worse. In one simulation in which the sample size was increased to 8,000 the coverage rate for the lognormal intervals dropped to 60%. This would seem to indicate that the underlying distribution of the expenditure data is not really lognormal but for smaller sample sizes the lognormal intervals cover at the desired rate because the intervals are really wide intervals.

(insert Figure I here)

Figure II represents coverage rates of the confidence intervals under stratified sampling. For stratified sampling the Empirical Bayes confidence intervals were used for the lognormal distribution. Note that for this kind of sampling, the coverage of the lognormal intervals does not start to deteriorate. Again, we see that the normal interval is not satisfactory at 1,000. The gamma interval coverage is good from small sample sizes. However, for stratified sampling the Empirical Bayes lognormal intervals appear to perform well, even for sample sizes up to 1,000.

(insert Figure II here)

The use of the gamma distribution clearly improved the coverage rate for the gamma intervals in the simulations. At this point, it cannot be ascertained from the simulations if the noted improvement is due to accounting for skewness of the data; if the improvement is a direct result of changing the nuisance parameter in the interval; or if it is possibly because of some combination of the two. In any event, the gamma intervals clearly outperformed the normal intervals in this simulation.

7. Conclusions and Directions for Future Research

The simulations indicate that normal confidence intervals can have quite poor coverage for samples less than 2,000, but the gamma distributional assumptions do improve on confidence interval coverage. Among the fixed subdomains the gamma interval had better coverage than the normal intervals in every case. The lognormal based intervals covered well for random subsample sizes less than 500, but coverage for these intervals deteriorated when subsample sizes approached 1,000. There were indications of serious coverage problems for the lognormal intervals. On the other hand, the gamma based intervals had good coverage for all random subsample sizes and good coverage in most subdomains. In addition, the width of the gamma intervals converged to the width of the normal intervals for larger random subsample sizes.

The simulations carried out as part of this research raise several interesting areas for possible future research: 1) estimation of the shape parameter of the gamma distribution, 2) determining the distribution of the gamma interval with an estimated shape parameter, 3) other problems beside skewness in the data and 4) the effect of multistage selection within strata on the coverage rates.

Estimation of the shape parameter for the gamma appeared to work satisfactorily in the simulations. However, the method of estimation is an ad hoc approach that needs improvement. The next research will focus on empirical Bayes construction of gamma intervals.

The exact small sample distribution of a gamma interval with an estimated shape parameter is a statistically interesting problem reminiscent of the Student t distribution. It is not clear how widely useful this distribution would be to the unsophisticated end user, but a better gamma interval would result. There may be the possibility of estimating this type of distribution with a bootstrap approach.

It can be conjectured that other distributional problems besides skewness are probably related to the semi-continuous nature of the data. In this study, the zero expenditures were eliminated and many analysts are only interested in estimating the nonzero expenditures. However, confidence intervals for expenditure data including the zero expenditures are also of interest so extending the work in this direction would be useful.

From the Bayesian perspective it would be interesting to base the Empirical Bayes confidence intervals on a gamma distribution.

From a survey practitioner's point of view, the complex sampling within strata may be the most interesting question. The current simulations only looked at simple random sampling within strata, but cluster sampling is known to have an impact on interval estimation as well as point estimates.

Acknowledgements
The authors would like to thank Steve Machlin, Lap-Ming Wun, Joel Cohen, and Steve Cohen for review and helpful comments. The views expressed in this paper are those of the authors and no official endorsement by the Department of Health and Human Services or the Agency for Healthcare Research and Quality is intended or should be inferred.

References

Berk and Monheit (2001) "The Concentration of Health Care Expenditures, Revisited," Health Affairs 20, no. 2.

Carlin, B.P. and Louis, T.A. (2000) *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd Ed, New York, Chapman & Hall/CRC

Cochran, W.G. (1977, p42). *Sampling Techniques.* New York: John Wiley and Sons

Gambino, J.G. (2003). "pps": Functions for PPS sampling. R package version 0.93.

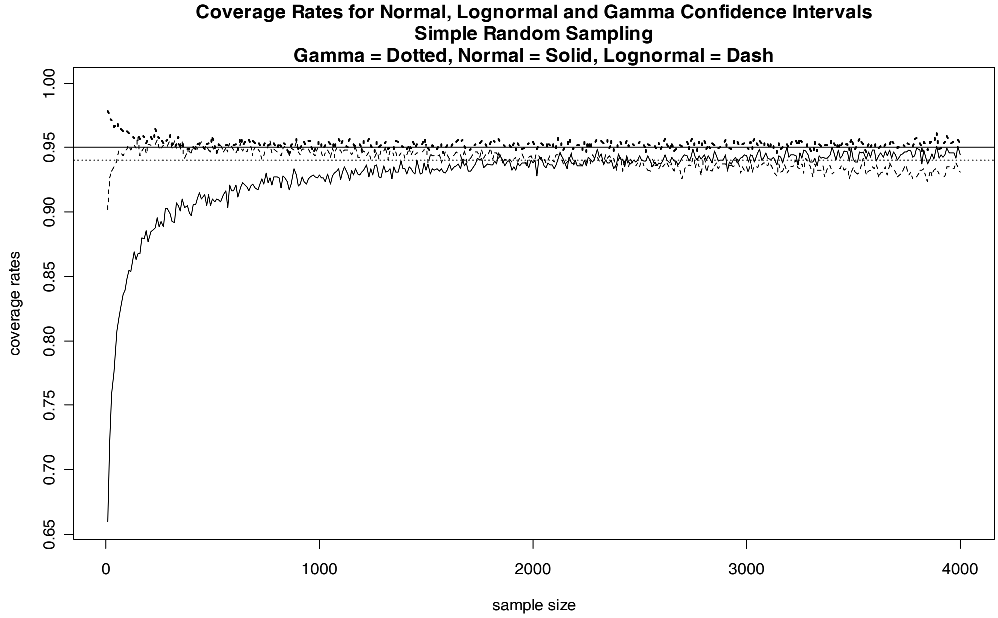
Land, C.E. (1972). An Evaluation of Approximate Confidence Interval Estimation Methods for Lognormal Means. *Annals of Mathematical Statistics*, 42, 1187-1205.

R Development Core Team (2005). "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Yu, W. and Machlin, S. (2004). Examination of Skewed Health Expenditure Data from the Medical Expenditure Panel Survey. *Proceedings of the Survey Research Methods Section, American Statistical Association*.

Zhou, X. and Gao, S. (1997). Confidence Intervals for the Log-Normal Mean. *Statistics in Medicine*, 16, 783-790.

Figure I



source of data: 1996-2001 MEPS

Figure II