

Sampling distribution

From Wikipedia, the free encyclopedia

In statistics, a **sampling distribution** or **finite-sample distribution** is the probability distribution of a given statistic based on a random sample. Sampling distributions are important in statistics because they provide a major simplification en route to statistical inference. More specifically, they allow analytical considerations to be based on the sampling distribution of a statistic, rather than on the joint probability distribution of all the individual sample values.

Contents

- 1 Introduction
- 2 Standard error
- 3 Examples
- 4 Statistical inference
- 5 References
- 6 External links

Introduction

The **sampling distribution** of a statistic is the distribution of that statistic, considered as a random variable, when derived from a random sample of size n . It may be considered as the distribution of the statistic for *all possible samples from the same population* of a given size. The sampling distribution depends on the underlying distribution of the population, the statistic being considered, the sampling procedure employed, and the sample size used. There is often considerable interest in whether the sampling distribution can be approximated by an asymptotic distribution, which corresponds to the limiting case either as the number of random samples of finite size, taken from an infinite population and used to produce the distribution, tends to infinity, or when just one equally-infinite-size "sample" is taken of that same population.

For example, consider a normal population with mean μ and variance σ^2 . Assume we repeatedly take samples of a given size from this population and calculate the arithmetic mean \bar{x} for each sample – this statistic is called the sample mean. Each sample has its own average value, and the distribution of these averages is called the "sampling distribution of the sample mean". This distribution is normal $\mathcal{N}(\mu, \sigma^2/n)$ (n is the sample size) since the underlying population is normal, although sampling distributions may also often be close to normal even when the population distribution is not (see central limit theorem). An alternative to the sample mean is the sample median. When calculated from the same population, it has a different sampling distribution to that of the mean and is generally not normal (but it may be close for large sample sizes).

The mean of a sample from a population having a normal distribution is an example of a simple statistic taken from one of the simplest statistical populations. For other statistics and other populations the formulas are more complicated, and often they don't exist in closed-form. In such cases the sampling distributions may be approximated through Monte-Carlo simulations^{[1][p. 2]}, bootstrap methods, or asymptotic distribution theory.

Standard error

The standard deviation of the sampling distribution of a statistic is referred to as the standard error of that quantity. For the case where the statistic is the sample mean, and samples are uncorrelated, the standard error is:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where σ is the standard deviation of the population distribution of that quantity and n is the sample size (number of items in the sample).

An important implication of this formula is that the sample size must be quadrupled (multiplied by 4) to achieve half (1/2) the measurement error. When designing statistical studies where cost is a factor, this may have a role in understanding cost–benefit tradeoffs.

Examples

Population	Statistic	Sampling distribution
Normal: $\mathcal{N}(\mu, \sigma^2)$	Sample mean \bar{X} from samples of size n	$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
Bernoulli: $\text{Bernoulli}(p)$	Sample proportion of "successful trials" \bar{X}	$n\bar{X} \sim \text{Binomial}(n, p)$
Two independent normal populations: $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$	Difference between sample means, $\bar{X}_1 - \bar{X}_2$	$\bar{X}_1 - \bar{X}_2 \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$
Any absolutely continuous distribution F with density f	Median $X_{(k)}$ from a sample of size $n = 2k - 1$, where sample is ordered $X_{(1)}$ to $X_{(n)}$	$f_{X_{(k)}}(x) = \frac{(2k - 1)!}{(k - 1)!^2} f(x) \left(F(x)(1 - F(x))\right)^{k - 1}$
Any distribution with distribution function F	Maximum $M = \max X_k$ from a random sample of size n	$F_M(x) = P(M \leq x) = \prod P(X_k \leq x) = (F(x))^n$

Statistical inference

In the theory of statistical inference, the idea of a sufficient statistic provides the basis of choosing a statistic (as a function of the sample data points) in such a way that no information is lost by replacing the full probabilistic description of the sample with the sampling distribution of the selected statistic.

In frequentist inference, for example in the development of a statistical hypothesis test or a confidence interval, the availability of the sampling distribution of a statistic (or an approximation to this in the form of an asymptotic distribution) can allow the ready formulation of such procedures, whereas the development of procedures starting from the joint distribution of the sample would be less straightforward.

In Bayesian inference, when the sampling distribution of a statistic is available, one can consider replacing the final outcome of such procedures, specifically the conditional distributions of any unknown quantities given the sample data, by the conditional distributions of any unknown quantities given selected sample statistics. Such a procedure would involve the sampling distribution of the statistics. The results would be identical provided the statistics chosen are jointly sufficient statistics.

References

1. Mooney, Christopher Z. (1999). *Monte Carlo simulation*. Thousand Oaks, Calif.: Sage. ISBN 9780803959439.
- Merberg, A. and S.J. Miller (2008). "The Sample Distribution of the Median". *Course Notes for Math 162: Mathematical Statistics*, on the web at http://web.williams.edu/Mathematics/sjmiller/public_html/BrownClasses/162/Handouts/MedianThm04.pdf, pgs 1–9.

External links

- Generate sampling distributions in Excel (<http://www.indiana.edu/~jkkteach/ExcelSampler/>)
- *Mathematica* demonstration showing the sampling distribution of various statistics (e.g. Σx^2) for a normal population (<http://demonstrations.wolfram.com/StatisticsAssociatedWithNormalSamples/>)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Sampling_distribution&oldid=691991987"

Categories: Statistical theory

-
- This page was last modified on 23 November 2015, at 12:55.
 - Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.