

A bit on sampling distributions

sampdis.tex Feb 28, 2003

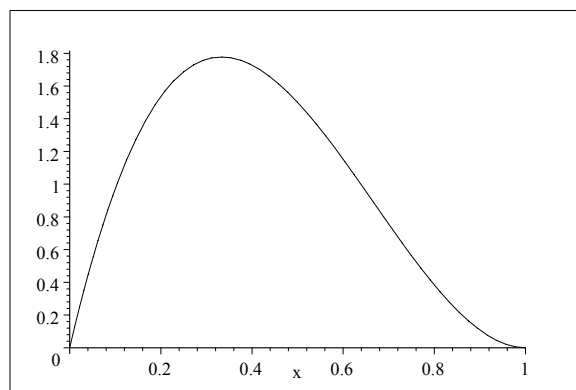
Note that \bar{X} is a random variable - it varies from sample to sample. so does any other function of (X_1, X_2, \dots, X_n) , such as $.5X_1 + .25X_2$, $\min(X_1, X_2, \dots, X_n)$, and X_3 .

Since statistics are random variables, they have density functions (in this case called sampling distributions because the realization of the statistic will vary across samples)

Understanding sampling distributions is the foundation of understanding econometrics. Every estimator (no matter how good or bad) has a sampling distribution. We choose one estimator over another on the basis of their sampling distributions. That is, we choose the one with the "better" sampling distribution.

Denote the density function for the sampling distribution of \bar{X} , $f_{\bar{X}}(\bar{x})$. In comparison, denote the sampling distribution of X_3 , $f_{X_3}(x_3)$. We have found the expected values (means) of these two sampling distributions/density functions

The mean of $f_{\bar{X}}(\bar{x})$, $E[\bar{X}]$, is μ_x , and the mean of $f_{X_3}(x_3)$, $E[X_3]$ is μ_x .



Possible sampling distribution of \bar{X}

We can also determine the $var[\bar{X}]$

$$\begin{aligned} var[\bar{X}] &= var\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \\ &= \frac{1}{n^2} var\left[\sum_{i=1}^n X_i\right] \text{ because } var[aX] = a^2 var[X] \\ &= \frac{1}{n^2} var[X_1 + X_2 + \dots + X_n] \\ &= \frac{1}{n^2} (var[X_1] + var[X_2] + \dots + var[X_n]) \end{aligned}$$

because $var[X_1 + X_2] = var[X_1] + var[X_2]$ if X_1 and X_2 are independent, which they are if it is a random sample.

$$= \frac{1}{n^2} (n\sigma_x^2) = \frac{\sigma_x^2}{n}$$

That is $var[\bar{X}] = \frac{\sigma_x^2}{n}$. It decreases, approaching zero, as n approaches ∞ .

Summarizing $E[\bar{X}] = \mu_x$ and $var[\bar{X}] = \frac{\sigma_x^2}{n} \forall f_X(x; \mu_x, \sigma_x^2)$, independent of the particular form of $f_X(x)$ if the sample is random.

Remember that we determined a bit ago, that X_3 was also an unbiased estimate of μ_x . What is $var[X_3]$ and how does it compare to $var[\bar{X}]$?

$$var[X_3] = \sigma_x^2$$

Compare this with $var[\bar{X}] = \frac{\sigma_x^2}{n}$. Both \bar{X} and X_3 are unbiased estimates of μ_x but we prefer the former to the latter because it has a much smaller variance. That is, we like its sampling distribution better because it is more concentrated around μ_x .

What to conclude? Ceteris, paribus, if two estimators are both unbiased. Go with the one that has the smaller sampling variance.

If one wanted to estimate μ_x with \bar{X} , would you prefer a sample size of 1, n , or $n + m$? Why? Ceteris paribus, more data/information is always better, assuming it is good data

Choose one or two other examples of statistics and derive the mean and variance of their sampling distributions.

For example, consider the statistic $\bar{X}/10$. What is its expectation? What is its variance and how does it relate to the variance of \bar{X} .

In your sampling distribution assignment you simulated a sampling distribution of the max X from samples of size N drawn from the normal distribution.

You found that the mean of the maximums increased as N increased, and the variance approached zero as the sample sizes increased.

MGB (sec 3.2 "Distribution on Minimum and Maximum) has some theorems on the distribution of max and min.

Theorem 5: If X_1, \dots, X_n are independent random variables and $Y_n = \max[X_1, \dots, X_n]$ then

$$F_{Y_n}(y) = \prod_{i=1}^n F_{X_i}(y)$$

And if X_1, \dots, X_n are independent and identically distributed with common cumulative distribution function $F_X(\cdot)$, then

$$F_{Y_n}(y) = [F_{X_i}(y)]^n$$

A corollary: if X_1, \dots, X_n are independent and identically distributed with common probability density function $f_X(\cdot)$ and cumulative density function $F_X(\cdot)$, then

$$f_{Y_n}(y) = n[F_{X_i}(y)]^{n-1}f_X(y)$$

How does this apply to your assignment about the sampling distribution of the max from a normal? The cdf for the max is the cdf for the normal raised to the power of the sample size.

A little intuition

What is the probability that some number y is at least as large as the largest X_i . It is

$$\Pr[X_1 \leq y; \dots; X_n \leq y]$$

If the X 's are independent then

$$\begin{aligned} & \Pr[X_1 \leq y; \dots; X_n \leq y] \\ &= \prod_{i=1}^n \Pr[X_i \leq y] = \prod_{i=1}^n F_{X_i}(y) \end{aligned}$$

If each of the X 's is an independent draw from the same distribution (e.g. a random sample) then

$$\begin{aligned} & \Pr[X_1 \leq y; \dots; X_n \leq y] \\ &= [F_{X_i}(y)]^n \end{aligned}$$

We just proved the above theorem. The corollary follows because it is always the case that $\frac{\partial F(x)}{\partial x} = f(x)$.

Deriving the sample distribution of the statistic, S

Since $f_{\bar{x}}(\bar{x})$ is a distribution, and since we can find the mean and variance of that distribution as a function of $f_X(x)$, can one determine the form of $f_{\bar{x}}(\bar{x})$ from knowledge of $f_X(x)$?

In theory, Yes, but it can be difficult. MGB have some examples on pp236-238. Note that just because $f_X(x)$ has a *Snerd* distribution does not mean that $f_{\bar{x}}(\bar{x})$ has a *Snerd* distribution. Most of the time it won't.

If, and only if, $X \sim N(\mu_x, \sigma_x^2)$, is $\bar{X} \sim N(\mu_x, \sigma_x^2/n)$
For details see MGB (pp241 and 246)

However, there is a theorem called the Central Limit Theorem (MGB 234 and G?) that implies that if x_1, x_2, \dots, x_n is a random sample

as $n \rightarrow \infty$ $f_{\bar{X}}(\bar{x}) \rightarrow$ a normal distribution

even if $f_X(x)$ is not normal.

This result will prove useful because

- often we cannot determine the form of $f_{\bar{X}}(\bar{x})$ even though we know $f_X(x)$
- often the normal will be a fairly good approximation for $f_{\bar{X}}(\bar{x})$ even when n is small.

In theory, one could assume some arbitrary density function $f_X(x)$ and some arbitrary statistic $s = s(x)$ and then theoretically derive the sampling distribution of s . However, this is often very difficult if not impossible.