1 The joint density of the sample: population, sample, and random sample

erm

Oct 26, 2010

In general, we assume the rv of interest, X, has some known distribution, $f_X(x)$ in the population of interest, but we do not known the values of the parameters in that distribution. Estimating their values is the goal.

- That is, we assume that in the population of interest, X has some **known** distribution
- This is a very strong assumption.
- Given that we know, by assumption, the distribution, the problem is to estimate the values of the parameters.

Many (most?) econometric problems are of this type.

We do estimation by taking a sample from the population of interest,and use information in the sample to estimate the values of the population parameters.

Hereafter, I will use the term *population* to refer to the *population of interest*. Sampling and estimation always starts by defining the population of interest.

If one wants to emphasis the parameters in $f_X(x)$, one might write it $f_X(x;\theta)$, where θ is the vector of parameters.

In a bit more detail: using data from a sample to make inferences about population parameters is an example of *induction*: generalizing from the particular to the general. Induction is an uncertain process: one sees a black cat and surmises that all cats are black. Better to determine the probability that all cats are black given that one has a sample of one cat and that cat is black.

Induction is made legitimate, or at least more legitimate, by attaching probabilities to one's generalizations: "my best estimate of μ is .32 and and given this estimate, there is a 95% probability that the interval .22 to .42 includes μ." You get the idea.

Induction needs to be contrasted with deduction: making predictions on the basis of assumptions and definitions (building theories). Given a set of defintions and assumptions a prediction does or does not logically follow - there is no uncertainty once one "buys" the definitions and assumptions.[1]

[1] When one proves something, for example that under certain assumptions OLS estimates are BLUE, one is doing deduction.

1.1 Samples and random samples

I will go crazy if you talk about sampling and samples without also talking about the population. One cannot talk about sampling absent a population, sampling from what?

The population is the set of elements (people, things, countries, firms, etc.) for which one wants estimates of the population parameters. Sometimes we make a distinction between the *population of interest* (the one we want to study) and the *sampled population* (the population we end up studying). When I use the word *population*, I mean the *population of interest*.

Definition: A sample is a subset of the population.

When we collect a sample we measure each element sampled on some metric which we can denote with the rv X. The rv variable X can be a vector or a scalar; for now assume it is a scalar.

Denote a sample of size n as $(x_1, x_2, ..., x_n)$ where x_j is the jth observation in the sample (MGB 223), remember that x_j can be a vector.

Note that a sample is a vector of random variables with some joint distribution, $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$. **This point is critical**.

In explanation, each variable in $f_{X_1,X_2,...,X_n}(X_1, X_2, ..., X_n)$ is a random variable; that is, observation j can take different values, so observation j is a rv. Denote this random variable X_j, and the specific value it takes x_j.

$f_{X_1,X_2,...,X_n}(X_1, X_2, ..., X_n)$ is therefore a joint density function for the n random variables in a sample.

Notationally, distinguish between $f_{X_1,X_2,...,X_n}(X_1, X_2, ..., X_n)$ and $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$. The latter refers to the joint density evaluated for a sample, the former represent the joint distribution function.

Note that in $f_{X_1,X_2,...,X_n}(X_1, X_2, ..., X_n)$, order is important, X_k is the k^{th} draw

Everything **said to here** applies to samples whether they are random samples or not random samples.

Ceteris paribus, We usually prefer our sample to be a random sample from our population, $f_X(x)$.[2]

[2] In more detail, random samples are preferred to non-random samples. The rub is that it

Definition: The sample $(x_1, x_2, ..., x_n)$ is a random sample from $f_X(x)$ if

$$f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) = f_X(x_1)f_X(x_2)...f_X(x_2)$$

where $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$ is the joint distribution of the sample (MGB 223 and G74).

In words, $(x_1, x_2, ..., x_n)$ is a random sample from $f_X(X)$ if each observation is an independent draw from $f_X(X)$.

Just to be clear, let me write out the above in a little more detail. The sample $(x_1, x_2, ..., x_n)$ is a random sample from $f_X(x)$ if

$$\begin{aligned} f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) &= f_{X_1}(x_1)f_{X_2}(x_2)...f_{Xn}(x_2) \\ &= f_X(x_1)f_X(x_2)...f_X(x_2) \text{ which in less detailed notation is} \\ &= f(x_1)f(x_2)...f(x_2) \end{aligned}$$

because

$$f_{X_i}(x_i) = f_X(x_i)$$

that is, each observation in the sample has the same distribution.

Often we say a sample is random if the observations in it are *independent, identically distributed* - I.I.A.

That is, a sample is random if each observation in the sample is independently drawn from the same (identical) distribution.

Said loosely, the sample is random if for each observation, each value of the rv in the population has an equal chance of appearing as the *jth* observation, and this is true for all j.

is often easier (read cheaper) to get a non-random sample, but if one's sample is nonrandom, estimation is more complicated. That said, we estimate stuff all the time with data from non-random samples, taking the non-randomness into account.

Give me an example of a nonrandom sample. Start by identifying the population you are sampling from.

Can one tell, by observation, whether a sample is a random sample?

Note that *random* does not mean *representative.*

However as n increases, the sample will likely become more representative of the population. (How does one judge representativeness?)

Because of sampling variation, samples differ. That is, any two random samples of size n from the same population are likely to not exhibit the same values of the n random variables $X_1, X_2, ..., X_n$.

Remember that the n random variables $X_1, X_2, ..., X_n$ have some joint density function

$$f_{X_1,X_2,...,X_n}(X_1, X_2, ..., X_n)$$

We call this the *distribution of sample.* Each sample is a draw from this distribution.

Consider a discrete rv that can takes the value of 1 with probability p, and the the value 0 with probablility $(1-p)$ - it has a Bernoulli distribution. Picture a random sample from this population with three observations; that is $n = 3$. If each is a draw from the same Bernoulli distribution, $f(x_1) = p^{x_1}(1-p)^{1-x_1}$, $f(x_2) = p^{x_2}(1-p)^{1-x_2}$ and $f(x_3) = p^{x_3}(1-p)^{1-x_3}$.

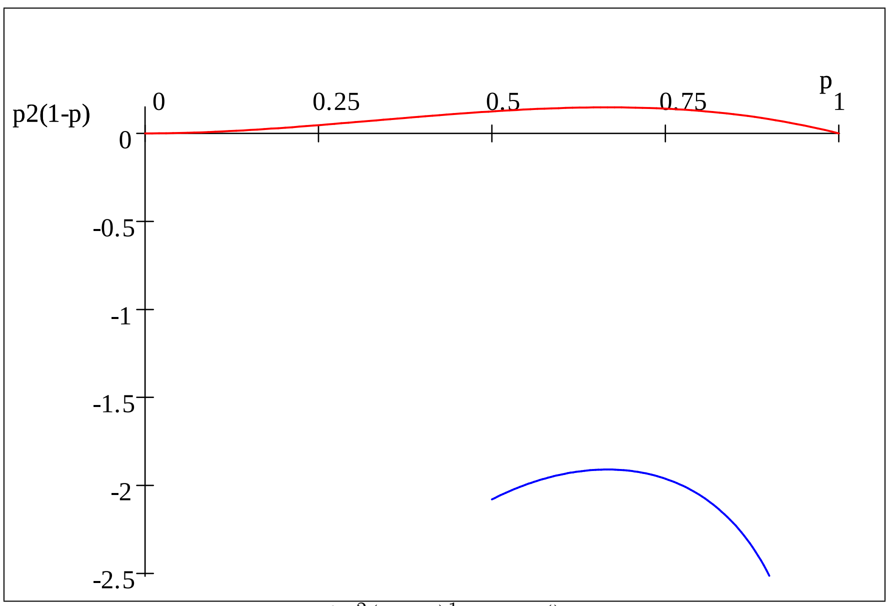
So,

$$\begin{aligned} f_{X_1,X_2,X_3}(x_1,x_2,x_3) &= f(x_1)f(x_2)f(x_3) \\ &= p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2}p^{x_3}(1-p)^{1-x_3} \\ &= p^{x_1+x_2+x_3}(1-p)^{3-x_1-x_2-x_3} \end{aligned}$$

is the joint density function for samples of three from this population.

Now assume your random sample consists of two ones and one zero. For fun, find the value of p, $\widehat{p}$, that maximizes the distribution of this sample. (Note that if $\widehat{p}$ maximizes $f_{X_1,X_2,X_3}(x_1,x_2,x_3)$, it also maximizes $\ln(f_{X_1,X_2,X_3}(x_1,x_2,x_3))$.

Given two ones and one zero $p^{x_1+x_2+x_3}(1-p)^{3-x_1-x_2-x_3} = p^2(1-p)^1$ Candidate(s) for extrema: $\left\{0, \frac{4}{27}\right\}$, at $\left\{[p=0], \left[p=\frac{2}{3}\right]\right\}$. The function is maximized at $\widehat{p} = \frac{2}{3}$. One finds the same thing by maximizing $2\ln p + \ln(1-p)$ Candidate(s) for extrema: $\left\{2\ln\frac{2}{3} - \ln 3\right\}$, at $\left\{\left[p=\frac{2}{3}\right]\right\}$. Either way works.



graph of $p^2(1-p)^1$ and ln() in blue

Keep in mind that the data was generated by some given, but unknown to you, p. Is the value of p that maximizes the distribution of the sample, an estimate of p? A good estimate of p?

Now assume a random sample of five observations, $(2, 2, 1, 0, 4)$ from the Poisson Distribution $f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0, 1, 2, 3, ...,\ \lambda > 0$.

Write down $f_{X_1,X_2,X_3}(x_1, x_2, x_3, x_{4,}x_5)$ for these five observations. For fun, find the value of λ that maximizes $f_{X_1,X_2,X_3}(x_1, x_2, x_3, x_{4,}x_5)$. Are you surprised. You think this is a good estimate?