1 A bit on sampling distributions

erm: sampdis.tex (draft) Nov 6, 2010

Note that a sample mean, $\bar{X}$, is a random variable - it varies from sample to sample.

Consider a sample of size n, $(X_1, X_2, ..., X_n)$.

Then consider a function of of the sample $f(X_1, X_2, ..., X_n)$. Such a function is called a statistic, so we might want to write it $s(X_1, X_2, ..., X_n)$

$\bar{X} = s(X_1, X_2, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ is one example

Other examples are $.5X_1 + .25X_2$, $\min(X_1, X_2, ..., X_n)$, and X_3.

All of these will take different values depending on specific sample drawn - they have *sampling variability.*

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 at the foundation of understanding econometrics.

For example, estimators (no matter how good or bad) have sampling distributions. **We typically choose one estimator over another on the basis of the properties of their sampling distributions.** That is, we choose the one with the "better" sampling distribution.

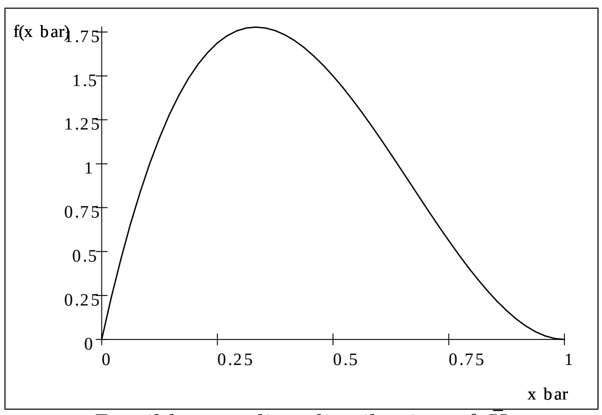
Consider a random variable X with density function $f(x)$ and mean μ_x.

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, $f_{X_3}(x_3)$, where X_3 is the third observation in the sample.

Two questions arise, what is the distribution/density function of a statistic, and what are its moments.

Think about 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 $\bar{X}$

Does it surprise you that they have the same means. Another question is do $\bar{X}$ and X_3 have the same density functions.

Consider $var[\bar{X}]$, when $\bar{X}$ is the mean of a random sample.

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

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.

$$var[\bar{X}] = \frac{1}{n^2}(n\sigma_x^2) = \frac{\sigma_x^2}{n}$$

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.[1]

[1] Looking ahead, the sampling distribution of $\bar{X}$ approaches a normal distribution, independent of the distribution of X, as the sample size increases. This is the central limit theorem. That is, if the sample size is large enough, we know that $\bar{X}$ is approximately normally distributed with mean μ_x and variance σ_x^2/n.

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 sampling 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 - one's estimate is likely to be closer to the population parameter.

If one wanted to estimate μ_x with $\bar{X}$, would you prefer a sample size of 1, n, or $n + m$? Why? Remember $var[\bar{X}] = \frac{\sigma_x^2}{n}$. Ceteris paribus, more data/information is always better, assuming it is good data. Increasing the sample size, improves finite-sample efficiency, somthing we have not defined, yet.

1.1 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 $\bar{X}$.

1.2 Deriving the sample distribution of a statistic, S

One way to derive/approximate the sampling distribution of a statistic is to simulate it. That is, draw a couple thousand random samples, calculate the statistic for each of the samples, and plot the distribution of the statistic across the samples.

This is always possible, no matter how complicated the statistic.

I want you to try this out in Mathematica. Specify a density function for a population. Don't choose some standard run-of-the-mill distribution.

Then choose a sampe size n, some specific number. Then specify some functional form for a statistic of this sample. Then choose some number of samples M. Draw M random samples, calculate the value of your statistic for each of your M samples, each of size n. Plot the sampling distributin of your statistic. Play around with different M, n and statistics. Maybe you could do this for a mixture distribution. That could be an assignment for one of the groups,

Or, in contrast to stimulationg the sampling distribution, one can attempt to theoretically derive the sampling distribution of the statistic $S = s(X_1, X_2, ..., X_n)$ based on $f_X(x)$ or $F_X(x)$, and the specification of $S = s(X_1, X_2, ..., X_n)$.

The process by which this would be accomplished is not so difficult to conceptualize, but often impossible to carry out in practice. More on this below. A lot of theoretical work in econometrics is derivations of sampling distributions of estimators.

1.2.1 The sampling distribution of $\bar{X}$

Note that if, and only if, $X\tilde{}N(\mu_x, \sigma_x^2)$, is $\bar{X}\tilde{}N(\mu_x, \sigma_x^2/n)$

For details see MGB (pp241 and 246)

However, there is a theorem called the Central Limit Theorem (MGB 234) that implies that if $x_1, x_2, ..., x_n$ is a random sample

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

even if $f_X(x)$ is not Normal.[2]

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 then theoretically derive the sampling distribution of $\bar{X}$. However, this is often very difficult, if not impossible.

[2] Note that this theorem does not imply that for some other statistic, S, that as $n \to \infty$ $f_S(s) \to$ a normal distribution

——————————————————————————————

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

The following is an example of deriving the distribution of a statistic.

Consider the rv $Y_n = \max[X_1, ...X_n]$. Y_n which is a statistic. The probability that some number m is at least as large as the largest X_i in a sample of size n is

$$\Pr\left[X_1 \leq m; ...; X_n \leq m\right]$$

Said differently, this is the probability that m is greater than or equal to all of the $X's$ in the sample.

If the X's are independent (for example, from a random sample) then

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

So we have determined, in general terms, $\Pr\left[X_1 \leq m; ...; X_n \leq m\right]$ if the $X's$ are from a random sample.

But note that $\Pr\left[X_1 \leq m; ...; X_n \leq m\right] = \Pr[y < m]$ where y is the largest value of X. So,[3]

$$\Pr[y < m] = F_{Y_n}(m) = \prod_{i=1}^{n} F_{X_i}(m)$$

And if $X_1, ...X_n$, already assumed independently distributed, are also **identically** distributed with common cumulative distribution function $F_X(.)$, then

$$\Pr[y < m] \equiv F_{Y_n}(m) = \left[F_X(m)\right]^n$$

We have just derived a sampling distribution for the statistic, the largest value of X in a sample of size n.

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

$$f_{Y_n}(y) = n\left[F_X(y)\right]^{n-1} f_X(y)$$

[3] Note that this formula works as long as the X_i are independent and allows for each of the X_i to have a different CDF, so is not specific to random samples.

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

How does this apply to the sampling distribution of the max of a random sample from a univariate normal?

The cdf for the max is the cdf for the normal raised to the power of the sample size.

1.2.3 More on deriving the sampling distribution of a statistic $S = s(X_1, X_2, ..., X_n)$

As we did above, one can theoretically derive the sampling distribution of the statistic $S = s(X_1, X_2, ..., X_n)$ based on $f_X(x)$ or $F_X(x)$, and the specification of $S = s(X_1, X_2, ..., X_n)$.

Before considering further the derivation of the distribution of a sta-

tistic, typically a function of a vector of random variables, consider deriving the the distribution of a single random variable, a simpler problem Derive $f_Y(y)$ knowing $y = g(x)$ and $F_X(x)$.

Consider first the derivation of the CDF of the sampling distribution of Y.

We can often determine $F_Y(m)$ from $F_X(x)$ by noting that $F_Y(m) = \Pr[y < m]$
$= \Pr[g(x) < m]$ because $y = g(x)$

In terms of the CDF
$F_Y(m) = F_X(g(x) < m) = \int\limits_{\{x:g(x)<m\}} f_X(x)dx$, the area under $f_X(x)$ consistent with $g(x) < m$.

So

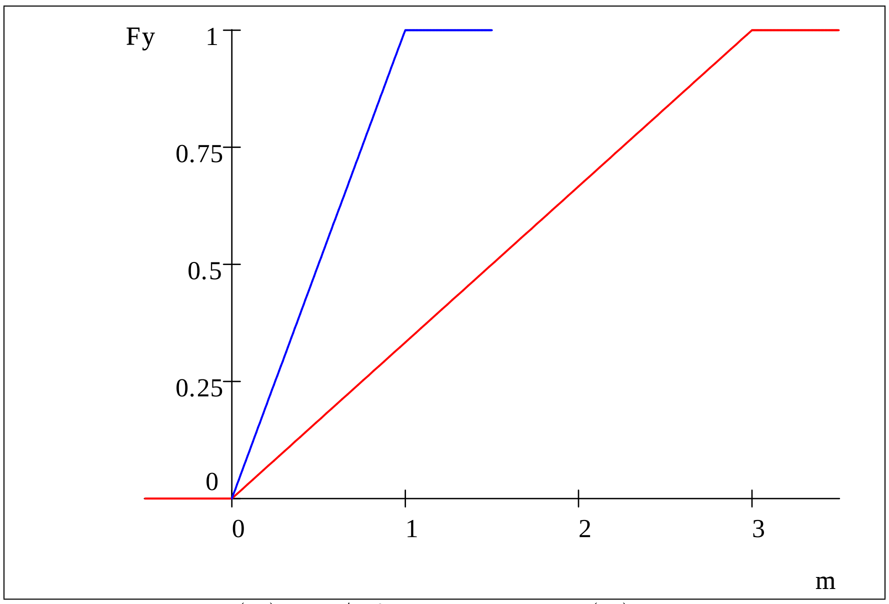
$$F_Y(m) \equiv \Pr[y < m] = \int\limits_{\{x:g(x)<m\}} f_X(x)dx$$

For example, if $y = g(x) = 3x$, an increasing function, then $x = g^{-1}(y) = \frac{1}{3}y$, and $F_Y(m) \equiv \Pr[y < m] = F_X(g(x) < m) = F_X(3x < m) = F_X(x < m/3)$, So

$$F_Y(m) = F_X(\tfrac{1}{3}m) = \int\limits_{\{x:3x<m\}} f_X(x)dx = \int\limits_{\{x:x<m/3\}} f_X(x)dx = \int\limits_{-\infty}^{m/3} f_X(x)dx$$

The answer, to here, is independent of the form of $f_X(x)$.

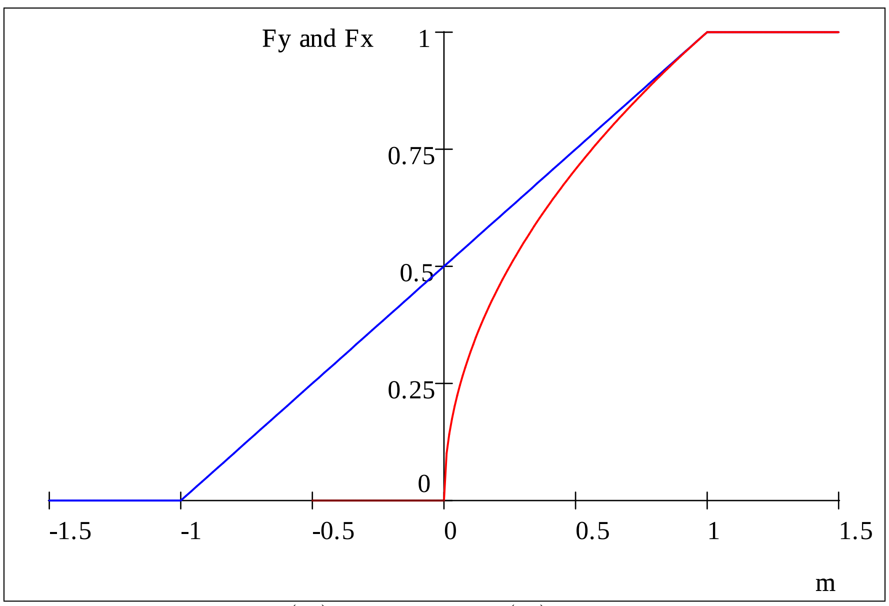
If, for example $f_X(x) = 1$ if $0 \leq x \leq 1$ and zero otherwise, $F_Y(m) = \int\limits_0^{m/3} 1dx = m/3$ for $0 \leq m \leq 3$, equals 0 if $m < 0$, and equals 1 if $m > 3$.



$F_Y(m) = m/3$ for $0 \leq m \leq 3$, $F_X(m)$ blue

Alternatively, assume $y = g(x) = x^2$. Note that since $x^2 \geq 0$, $\Pr[x^2 < m] = 0$ if $m < 0$. $F_Y(m) = \Pr[x^2 < m] = \Pr[-\sqrt{m} \leq x \leq \sqrt{m}] = F_X(\sqrt{m}) - F_X(-\sqrt{m}) = 2\int\limits_0^{\sqrt{m}} f_X(x)dx$, $m > 0$, and zero otherwise.[4]

And if, for example, $f_X(x) = .5$ if $-1 \leq x \leq 1$ and zero otherwise, $F_Y(m) = 2\int\limits_0^{\sqrt{m}} .5dx = \sqrt{m}$.[5] Or, in terms of the corresponding cdf $F_X(x) = .5 + .5x$ if $-1 \leq x \leq 1$, $F_X(\sqrt{m}) - F_X(-\sqrt{m}) = (.5 + .5\sqrt{m}) - (.5 + .5(-\sqrt{m})) = \sqrt{m}$. So, for example $\Pr[y < .5] = \sqrt{.5} = 0.707\,11$.



$F_Y(m)$ in red and $F_X(m)$ in blue

---

[4] Because, since $m > 0$ $F_X(\sqrt{m}) - F_X(-\sqrt{m}) = \int\limits_{-\infty}^{0} f_X(x)dx + \int\limits_0^{\sqrt{m}} f_X(x)dx$

$- \int\limits_{-\infty}^{0} f_X(x)dx - \int\limits_0^{-\sqrt{m}} f_X(x)dx$

$= \int\limits_0^{\sqrt{m}} f_X(x)dx - \int\limits_0^{-\sqrt{m}} f_X(x)dx = 2\int\limits_0^{\sqrt{m}} f_X(x)dx$

[5] Implying $f_Y(m) = \frac{d\sqrt{m}}{dm} = \frac{1}{2\sqrt{m}}$.

Theorem: In those cases where $y = g(x)$ is strictly increasing (or strictly decreasing) in x, so that the inverse function, $g^{-1}(y)$ exists, it is possible to show that

$$\begin{aligned} F_Y(m) &= \Pr[g(x) < m] \\ &= \begin{cases} \Pr[x < g^{-1}(m)] = F_X(g^{-1}(m)) & if \quad g(x) \uparrow x \\ \Pr[x > g^{-1}(m)] = 1 - F_X(g^{-1}(m)) & if \quad g(x) \downarrow x \end{cases} \end{aligned}$$

Demonstrating with a the simpliest example, $y_1 = g_1(x) = x$, where we know in advance that $F_{Y_1}(m) = F_X(m)$. Proceeding with the above steps, since $g_1(x)$ is increasing in x $F_{Y_1}(m) = \Pr[g_1(x) < m] = \Pr[x < g_1^{-1}(m)] = \Pr[x < m] = F_X(m)$, which is correct.

But now consider $y_2 = g_2(x) = -x$, where $g_2(x)$ is decreasing in x. In this simple case, $F_{Y_2}(m) = \Pr[g_2(x) < m] = \Pr[-x < m] = \Pr[x > -m] = 1 - F_X(-m)$, which is what one get if one applies the above equation for a decreasing function. That is, $F_{Y_2}(m) = \Pr[g_2(x) < m] = \Pr[x > g_2^{-1}(m)] = \Pr[x > -m] = 1 - F_X(-m)$

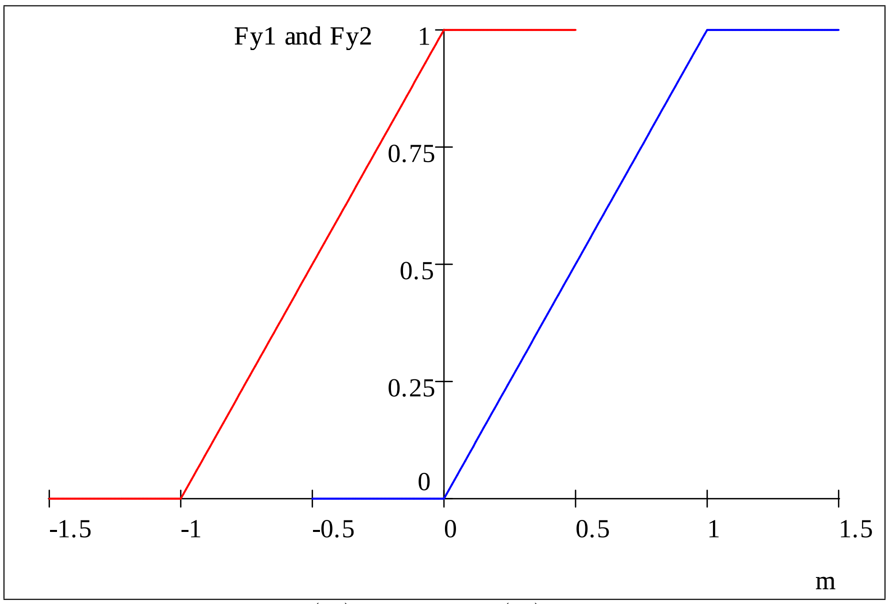
Demonstrating that the theorem is correct, at least, for $g(x) = x$ and $g(x) = -x$.

If, for example $f_X(x) = 1$ if $0 \leq x \leq 1$ and zero otherwise$\Longleftrightarrow F_X(x) = x$ if $0 \leq x \leq 1$ and zero otherwise,

$$F_{Y_1}(m) = F_X(m) = \begin{cases} 0 & if & m < 0 \\ m & if & 0 \leq m \leq 1 \\ 1 & if & m > 1 \end{cases}$$

and

$$F_{Y_2}(m) = 1 - F_X(-m) = \begin{cases} 0 & if & m < -1 \\ 1+m & if & -1 \leq m \leq 0 \\ 1 & if & m > 0 \end{cases}$$

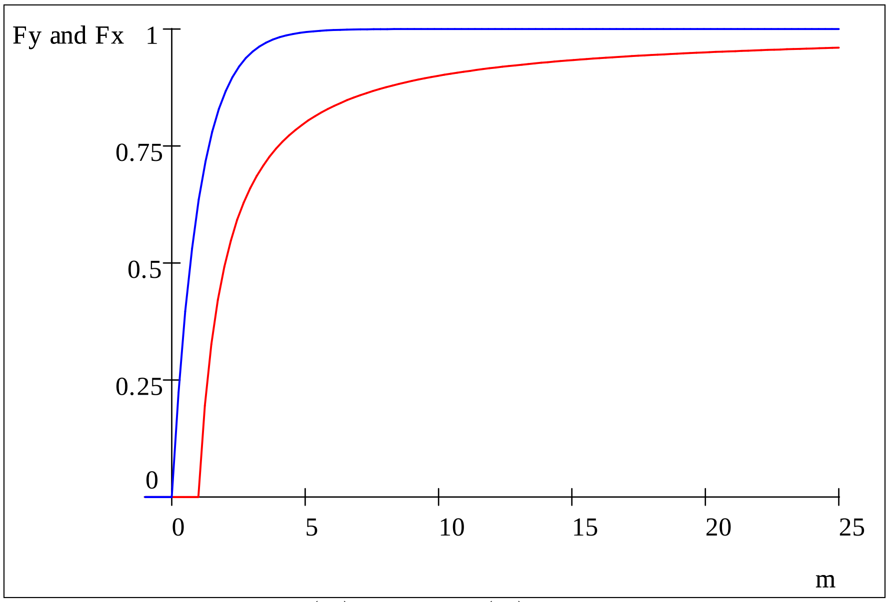


$F_{Y_1}(m)$ in blue, $F_{Y_2}(m)$ in red

Another example: if $y = g(x) = e^x$, $x = g^{-1}(y) = \ln y$,

$$F_Y(m) = \begin{cases} 0 & if \quad m < 0 \\ F_X(\ln m) & if \quad m \geq 0 \end{cases}$$

So, if, for example, $F_X(x) = 1 - e^{\frac{-x}{\mu}}$, $x \geq 0$, $F_Y(m) = 1 - e^{\frac{-\ln(m)}{\mu}} = 1 - \frac{1}{m^{\frac{1}{\mu}}}$, $m \geq 1$, and zero otherwise.[6] Graphing this for $\mu = 1$



$F_Y(m)$ in red, $F_X(m)$ in blue

[6]Note that $F_Y(m) = 0$ for $0 \leq m < 1$ even though the chosen $f_X(m)$ has positive density in this range.

Using $f_X(x) = \frac{dF_X(x)}{dx}$ and the above theorem, one can derive $f_Y(m)$. Consider first $\frac{dg(x)}{dx} > 0$, strictly increasing in x. In which case

$$\begin{aligned}
f_Y(m) &= \frac{dF_X(x)}{dx} = \frac{dF_X(g^{-1}(m))}{d(g^{-1}(m))}\frac{d(g^{-1}(m))}{dm} \\
&= f_X(g^{-1}(m))\frac{d(g^{-1}(m))}{dm} \\
&= f_X(g^{-1}(m))\left|\frac{d(g^{-1}(m))}{dm}\right|
\end{aligned}$$
[7]

Alternatively, if $\frac{dg(x)}{dx} < 0$

$$\begin{aligned}
f_Y(m) &= \frac{d\left[1 - F_X(x)\right]F_X(x)}{dx} = \frac{d\left[1 - F_X(g^{-1}(m))\right]}{d(g^{-1}(m))}\frac{d(g^{-1}(m))}{dm} \\
&= -f_X(g^{-1}(m))\frac{d(g^{-1}(m))}{dm} \\
&= f_X(g^{-1}(m))\left|\frac{d(g^{-1}(m))}{dm}\right|
\end{aligned}$$
[8]

So, putting these together

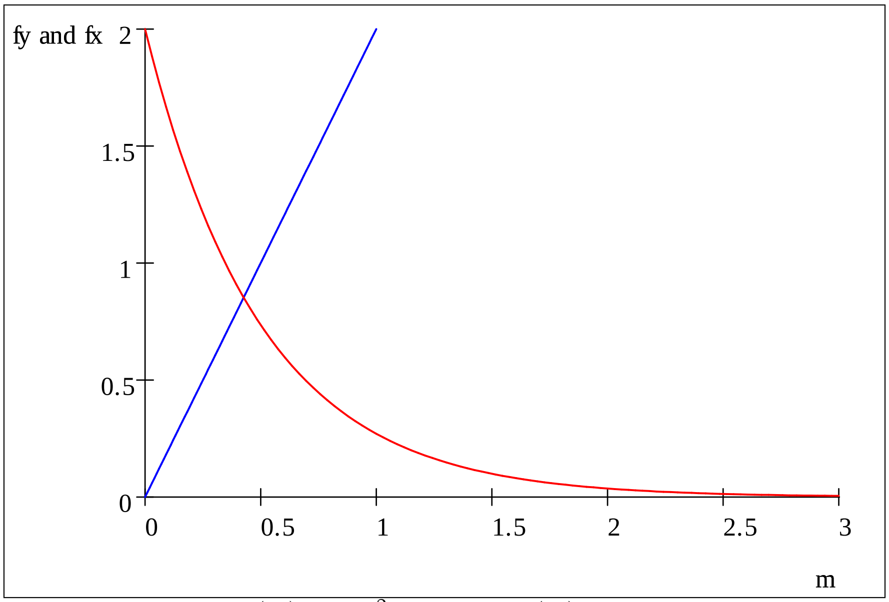
$$f_Y(m) = f_X(g^{-1}(m))\left|\frac{d(g^{-1}(m))}{dm}\right|$$

if $y = g(x)$ is strictly increasing (or decreasing) in x.

Example: Assume $y = -\ln x$, implying $g^{-1}(m) = e^{-m}$ and $\frac{d(g^{-1}(m))}{dm} = -e^{-m}$. And assume $f_X(x) = ax^{a-1}$, $0 \leq x \leq 1$ and zero otherwise, $a > 0$ (a restricted Beta distribution). In which case,

$$\begin{aligned} f_Y(m) &= f_X(g^{-1}(m)) \left| \frac{d(g^{-1}(m))}{dm} \right| \\ &= a(e^{-m})^{a-1} e^{-m} = ae^{-am} \end{aligned}$$

Graphing this for $a = 2$



$f_Y(m) = 2e^{-2m}$ in red, $f_X(m)$ in blue

$f_Y(m) = ae^{-am}$ is the exponential distribution.

In turns out that we can often use $f_Y(m) = f_X(g^{-1}(m))\left|\frac{d(g^{-1}(m))}{dm}\right|$ even if $g(x)$ is not strictly increasing or decreasing in x.

Consider a simple case where there are two segment: $x < 0$ and $x \geq 0$ where in one segment $\frac{dg(x)}{dx} > 0$ and in the other $\frac{dg(x)}{dx} < 0$. Let $g_{x<0}(x)$ refer to the function when $x < 0$ and $g_{x\geq 0}(x)$ refer to the function where $x \geq 0$. In this case (why?)

$$f_Y(m) = f_X(g_{x<0}^{-1}(m))\left|\frac{d(g_{x<0}^{-1}(m))}{dm}\right| + f_X(g_{x\geq 0}^{-1}(m))\left|\frac{d(g_{x\geq 0}^{-1}(m))}{dm}\right|$$

For example, consider our earlier example $y = g(x) = x^2$. It is strictly decreasing for $x < 0$ and increasing for $x > 0$. So for $x \geq 0$, $g_{x\geq 0}^{-1}(m) = \sqrt{m}$, $m > 0$, and for $x < 0$, $g_{x<0}^{-1}(m) = -\sqrt{m}$. So, $\frac{d(g_{x\geq 0}^{-1}(m))}{dm} = \frac{d\sqrt{m}}{dm} = \frac{1}{2\sqrt{m}}$ and $\frac{d(g_{x<0}^{-1}(m))}{dm} = \frac{d(-\sqrt{m})}{dm} = -\frac{1}{2\sqrt{m}}$. In which case,

$$\begin{aligned} f_Y(m) &= f_X(-\sqrt{m})\frac{1}{2\sqrt{m}} + f_X(\sqrt{m})\frac{1}{2\sqrt{m}} \\ &= \frac{1}{2\sqrt{m}}\left[f_X(-\sqrt{m}) + f_X(\sqrt{m})\right] \end{aligned}$$

for $m > 0$

And if, for example, $f_X(x) = .5$ if $-1 \leq x \leq 1$ and zero otherwise,

$$\begin{aligned} f_Y(m) &= \frac{1}{2\sqrt{m}}\left[f_X(-\sqrt{m}) + f_X(\sqrt{m})\right] \\ &= \frac{1}{2\sqrt{m}}[.5 + .5] = \frac{1}{2\sqrt{m}} \end{aligned}$$

if $m > 0$ and zero otherwise. This is same answer we got above where we used another technique to find the distribution of x^2 when $f_X(x) = .5$.

So, to find the sampling distribution of $S = s(X_1, X_2, ..., X_n)$ one needs to "vectorize" this technique. What do we know?

$F_S(m) \equiv \Pr[s < m] = \Pr[\, s(X_1, X_2, ..., X_n) < m]$.

The right-hand term is just a function of $s(X_1, X_2, ..., X_n)$ and the joint density $f_X(x_1, x_2, ..., x_n)$, so one should, in theory, be able to figure out this probability.

One can think of it as an integration problem:

$$F_S(m) \equiv \Pr[\, s(X_1, X_2, ..., X_n) < m] = \underset{s(x_1,x_2,...x_n)<m}{\int \cdots \int} f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) dx_1 dx_2 ... d_{x_n}$$

Doing this integration can be tough. This technique for finding the CDF for $s(X_1, X_2, ..., X_n)$ is called the *cummulative-distribution-function technique.*[9]

Consider a very simple example. Assume two random variables, and $S = s(X_1, X_2) = X_1 + X_2$, so we are looking for the cdf of the sum of two random variables:

$$\begin{aligned} F_Y(m) &= \Pr[y < m] = \Pr[\, X_1 + X_2 < m] \\ &= \iint_{x_1+x_2<m} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{m-x_1} f_{X_1,X_2}(x_1, x_2) dx_2 \right] dx_1 \end{aligned}$$

As x_1 goes from minus to plus infinity, x_2 goes from minus infinity to $m - x_1$, maintaining the restriction that $X_1 + X_2 < m$.

Can you come up with another example? An unoccupied group could come up with some good review question have to do with the derivation of a statistics sampling distribution.

There are other techniques, for finding the density of $s(X_1, X_2, ..., X_n)$: one is called the *moment-generating-function technique* - you can read about it in MGB. MGB devote most of a chapter to techniques for finding the distribution of $s(X_1, X_2, ..., X_n)$.

[9] Remember that for a specific statistic, $s(X_1, X_2, ..., X_n)$, and a specific density, $f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n)$, one can fall back to simulation.

Remember that if one only needs to know $E[S]$ and σ_S^2 one does not need to derive the complete distribution of S One can use the standard method of calculating the expected value of a rv which is a function of rv's

$$E[S] = \int_{-\infty}^{+\infty} \cdots \int s(x_1, x_2, ..., x_n) f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) dx_1 dx_2 ... d_{x_n}$$

$$E[S^2] = \int_{-\infty}^{+\infty} \cdots \int (s(x_1, x_2, ..., x_n))^2 f_{X_1,X_2,...,X_n}(x_1, x_2, ..., x_n) dx_1 dx_2 ... d_{x_n}$$

and

$$\sigma_S^2 = E[S^2] - E[S]^2$$

1.3 Consider one more sampling distribution

Consider the model $y_i = \beta x_i + \varepsilon_i$ where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. For each given x_i , one takes a random sample of (y_i, x_i) pairs and produces an estimate of β, $\widehat{\beta}$. $\widehat{\beta}$ will obviously have a sampling distribution. A big part of econometrics is determing the sampling distribution of this and other parameter estimators.

Could you now simulate such a sampling distribution? Sure, and at some point, I will ask you to do this.