



Guía Metodológica 2

La elucidación de las distribuciones de probabilidad

(documento interno del proyecto)

Autor: Gabriel Ángel Lacelli (lacelli.gabriel@inta.gob.ar)

Proyecto Disciplinario PDI065 “Gestión Integral del Riesgo Agropecuario”

Coordinadora: Silvina M. Cabrini (EEA Pergamino)

Producto 3. Indicadores de riesgo global para sistemas productivos en diferentes regiones.

En este producto se trabajará a nivel de unidad productiva con la simulación de resultados económicos. Los riesgos climáticos y de mercados se consideran utilizando precios y niveles de productividad en modelos de simulación estocástica.

Responsable del Producto 3. Gabriel A. Lacelli (EEA Reconquista)



En esta segunda guía trataremos de revisar algunos caminos que nos permitan producir las distribuciones de probabilidades de aquellas variables seleccionadas como críticas y cuyos valores deben ser aleatorizados.

Trabajar sólo con las variables críticas

Desde un enfoque sistémico es razonable pensar que no todas las variables que actúan en un proceso tienen igual categoría o importancia en la definición de los resultados. Precisamente los modelos tienen entre sus finalidades, la de simplificar la realidad que representan para hacerla más comprensible y facilitar los procesos de solución, ya sea por uso de algoritmos, métodos heurísticos o de simulación. En principio podemos reconocer diferentes grados de conocimiento del valor que asumirán las variables, a saber:

Grado de conocimiento	Grado de Influencia	Ejemplo
Variables conocidas	Modificables	<i>Variedades o híbridos de la semilla; nivel de fertilización</i>
	No modificables	<i>Tipo de suelo; genética del rodeo; precio de los insumos (no siempre)</i>
Variables desconocidas	No modificables o poco influenciados	<i>Cantidad y distribución de las lluvias, precio de los productos, rendimientos de las actividades</i>

En la aproximación que plantea esta guía, focalizaremos solamente sobre dos variables: **el precio de los productos y los rendimientos**. Ello no implica desconocer la importancia de otras, sino que es un intento por simplificar el tratamiento inicial de tan compleja y significativa temática.

De manera que al encarar la tarea de simular asumiremos en condición de certeza los valores del resto de las variables -las que en consecuencia pierden su condición de tal y serán tratadas como parámetros-, por lo que resta determinar el valor que otorgaremos o tendrán los precios y rendimientos de los productos que el sistema produce y que constituyen su ingreso.

Las variables económicas

En particular el tratamiento de las variables económicas requiere dos consideraciones previas. En primer lugar debe aclararse que cuando se opera en un contexto de estabilidad de precios pierde relevancia la unidad o la moneda en que se expresen las variables económicas. Sin embargo en un contexto inflacionario esto es de fundamental importancia, ya que la permanente modificación de los precios de los bienes y servicios así como de los precios relativos entre ellos, pueden introducir errores de magnitud en las decisiones y en la medición de los resultados.



En segundo lugar es importante definir la plaza sobre la que se estimarán los precios, ya que los costos variables asociados (gastos de fletes y comercialización), variarán en función que se trate de ventas en chacra, en mercados locales (por ej. en cooperativas, acopiadores locales, remate-feria, etc.) o en mercados de concentración referenciales más distantes, como pueden ser el complejo oleaginoso-cerealero de Rosario o el mercado de hacienda de Liniers.

Metodologías de Estimación

Metodológicamente el tipo de análisis a realizar tiene que ver con la disponibilidad de datos y la existencia de mercados de futuros. La existencia de datos históricos permite realizar análisis conocidos como **objetivos**, ya que la predicción se basa en información que realmente ocurrió. Por ejemplo, es común contar con datos de los rendimientos ocurridos a lo largo de una serie de años para un determinado cultivo en la zona, o los precios del producto para el último quinquenio o la última década. En este caso se trata de observaciones que realmente ocurrieron, por eso a los análisis sobre este tipo de datos se los llama “objetivos”.

Cuando no se dispone de tal información, ya sea por no hallarse disponible o bien porque se trata de una actividad nueva que no registra antecedentes zonales o bien porque se juzgan insuficiente a los análisis sobre datos históricos (ya que hay quienes consideran que el futuro no siempre es una “prolongación” o reiteración del pasado), se puede apelar a métodos **subjetivos**, los que se apoyan en estimaciones que los decisores realizan sobre un valor que ocurrirá en el futuro.

Ambos procedimientos (objetivos y subjetivos) presentan características que los hacen más o menos ventajosos y pertinentes, el detalle de las cuales escapa al alcance de esta guía. Se presentarán herramientas de análisis tanto de uno como de otro método.

Métodos apoyados en datos históricos

Las fuentes de datos históricos son las estadísticas que pueden hallarse en publicaciones de organismos oficiales (ministerios e institutos de estadísticas provinciales, MAGyP de la Nación, INDEC, BCRA) o privados (cooperativas, Bolsa de Comercio de Rosario, Anales de la Sociedad Rural Argentina, etc.), sitios web y revistas especializadas (FyO, Agromercado, Márgenes Agropecuarios, Precios CREA, entre otras)

También puede recurrirse a información de establecimientos de la zona (es lo ideal, pero lo menos común, por falta de registros). También, aunque con recaudos, de las propias producciones en nuestras EEA.

A partir de la serie de datos (también llamadas observaciones o distribución de una variable) pueden realizarse las siguientes operaciones.

1. Calcular los descriptores estadísticos. Promedio o media, moda, máximo y mínimo, varianza, desvío estándar y coeficiente de variación.



2. En lugar de trabajar con valores extremos correspondientes a un solo valor (máximos y mínimos absolutos), calcular el promedio de una parte o fractil de la distribución (ej. la media del decil superior, del 20% inferior, etc.). Sería similar a truncar la distribución.
3. Realizar histogramas que muestren la distribución de los datos, los que luego pueden transformarse en valores relativos o de probabilidad.
4. Realizar correcciones de las series originales. Este procedimiento es aplicable cuando disponemos de una serie larga de datos, pero que en los últimos años se ha producido alguna ruptura o cambio estructural.

Métodos apoyados en estimaciones subjetivas

En estos métodos los datos no son obtenidos a partir de observaciones históricas, sino a partir de la apreciación o expectativas que sobre los valores futuros de las variables analizadas, tengan los decisores o grupos de informantes calificadas.

Se trata de metodologías que apelan a rescatar la experiencia que sobre la ocurrencia futura de un precio o un rendimiento posea una persona o un grupo, que puedan ser calificados como idónea/o o experta/o en la temática bajo tratamiento.

En general se dividen en métodos paramétricos y no paramétricos, requiriéndose para los primeros la determinación de parámetros asociados a alguna distribución de probabilidades; en este sentido las distribuciones normales o triangulares son la de uso más difundido por presentar algunas ventajas prácticas. Los métodos no paramétricos intentan reconstruir la distribución de probabilidad a partir del conocimiento heurístico del decisor.

Mercados de Futuros

Un caso especial es el que constituyen los mercados de futuros, ya que permiten anticipar las expectativas que un conjunto de operadores sostiene en determinado momento presente sobre el valor futuro (precio) que poseerá una mercancía en el momento de ser comercializada. En nuestro país estos mercados existen para algunos cereales (trigo, maíz y sorgo) y oleaginosos (soja y girasol). También existe un valor referencial para el precio de la carne, a partir de índices novillo en el mercado de Liniers y en el Rofex.

¿Cómo vamos a trabajar en concreto?

Como utilizaremos algún software para hacer las distribuciones (*Simulación4.0* y/o *SimulAr*, más adelante, si es posible, *R* u otros más sofisticados), y los mismos ya traen las diferentes distribuciones tanto continuas como discretas, nuestra tarea consiste en proveerles la siguiente información: a qué tipo de distribución se ajustan nuestras variables y los parámetros que definen a dicha distribución, necesarios para que el software haga la simulación.



Entonces, tendremos dos situaciones de arranque. Puede ser que **dispongamos de una cantidad de datos suficientes y considerados adecuados para que estudiemos su naturaleza probabilística**, es decir encontrar la forma que mejor se ajusta a la distribución de datos. Luego, obtendremos los parámetros que definen o caracterizan la distribución, los cuales varían de acuerdo al tipo de distribución y son propios de cada una (por ejemplo: *media* y *desvío* en una Normal; *máximo*, *modal* y *mínimo* en una Triangular; *máximo* y *mínimo* en una Uniforme).

La segunda situación es que **no dispongamos de datos o sean pocos** (como un criterio práctico, menos de 10 datos podrían ser considerados insuficientes). La no disponibilidad de datos es bastante común, por lo menos en lo que hace a rendimientos de las actividades, así que debemos recurrir a métodos subjetivos para construir nuestras distribuciones. En principio, podemos suponer razonablemente que los rendimientos pueden distribuirse en forma normal o triangular, entonces lo que necesitamos es determinar los parámetros, ya sea por nuestro propio criterio o consultando a expertos (compañeros investigadores, extensionistas, colegas de la privada, productores). En la consulta debemos tener la precaución de proveer a la persona entrevistada, los datos generales del sistema, la localización del mismo y la tecnología de producción definida (una cosa es el abanico de rendimientos esperados si se fertiliza o no, si es en secano o riego, etc.).

Ejemplo de disponibilidad de datos

Tenemos una serie de precios de soja, maíz y trigo en Rosario, para el período 2002 a 2020, expresados en dólares. Si bien hubo cambios de políticas aplicadas durante el período analizado que afectaron los precios internos de los granos, nos permitimos asumir que en los tres casos, son datos homogéneos y comparables.

Consideramos solamente los precios de los principales meses de cosecha y venta, aunque esto cambia de una zona a otra del país y, además, podría ser variable de acuerdo a la estrategia de comercialización del sistema. Para soja tomamos abril, mayo, junio y julio; para maíz marzo a julio y para trigo noviembre a enero.

De esta manera, los resultados son los siguientes:

	Maíz	Soja	Trigo
Cantidad datos (n)	79	73	43
Promedio	124,4	241	141,5
Desvío	40,5	62,2	36,0
Máximo	201,1	384,9	222,0
Mínimo	64,1	121,5	82,8
Ajuste por software SimulAr			
Tipo de Distribución	Uniforme	Normal	Normal
Parámetros	Máx: 201,1	Media: 241	Media: 141,5
	Mín: 64,1	Desvío: 62,2	Desvío: 36
Ajuste por software Simulación4.0			
Distribución	Chi2	Chi2	Chi2
Parámetros	GL: 117	GL: 229	GL:133



Como puede observarse el SimulAr presenta una mayor versatilidad, ya que nos permite probar la bondad del ajuste de diferentes distribuciones y ver, consecuentemente, el mayor, menor o nulo grado de ajuste. El Simulación4.0 en este aspecto es más “duro”, ya que nos da un solo resultado y no nos deja “jugar”. Al hacer simulaciones aplicando las distribuciones sugeridas por uno y otro soft, siempre las del Simulación4.0 (en las tres series arrojó chi cuadrado), resultaron menos variables, con medias similares, algo menores que para las formas sugeridas por el SimulAr, pero máximos más chicos y mínimos mayores. Es decir, que introducen menos variabilidad en el análisis, alejándose en estos tres casos de una representación más cabal de lo que realmente ocurre con los precios. Conclusión mía: el SimulAr presenta mayor robustez para este aspecto, pero debemos reservar la decisión de cuál distribución utilizar analizando cada caso en particular.

Ejemplo cuando no disponemos de datos

Lo habitual es que para las variables que necesitamos simular, no existan datos suficientes y homogéneos sobre los cuáles estudiar el tipo de distribución y los parámetros que la definan. Tal vez para los precios en general haya datos, pero para rendimientos es más complicado. Existen datos pero con ciertos niveles de agregación que aplanan la variabilidad (generalmente a nivel departamental o provinciales). También hay que tener en cuenta los cambios tecnológicos, que hacen que los rendimientos obtenidos hoy no sean comparables con los obtenidos hace 5 o 10 años.

De manera que nos enfrentamos con dos temas. Por un lado determinar, saber, averiguar, cuál es el tipo de distribución que siguen los datos. Una vez resuelto este punto, restaría saber cuáles son los parámetros a los que responde la distribución.

Lo que estamos haciendo es asumir distribución triangular y, asignarle a la misma, valores para los parámetros que la definen (máximo, modal y mínimo). Ejemplo, si nuestro sistema incluye la actividad de cría vacuna, aleatorizamos la variable destete y decimos que la misma se distribuye de acuerdo a una triangular con máximo 65%, mínimo 58% y 63% como valor más frecuente ¿de dónde salen esos valores? Seguramente de consultas que hacemos a informantes expertos ¿de dónde sale que la variable destete se distribuye en forma triangular? Bueno, esto ya es más intuitivo, pero parece más razonable pensar que el destete ajusta mejor a esa distribución que a una weibull, una gamma o una exponencial. Pero en todo caso es un asunto que no resolveremos en esta guía y que debería ser motivo de una línea de investigación más específica (tal vez haya alguien con inclinaciones hacia la matemática y la estadística que se dedique a esto).

Sugerencia: mientras no tengamos una propuesta superadora (o elementos que justifiquen otra decisión), en el caso de variables para las que no dispongamos un data set robusto para determinar la forma de su distribución, asumiremos la distribución triangular y sus parámetros los definiremos recurriendo al aporte de informantes expertos.