



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Microsimulación

Docente: Daniel Ciganda

9^{na} Clase

5 de Noviembre de 2025

- En las próximas clases nos enfocaremos en los **modelos de microsimulación (MS) dinámicos, a gran escala y multipropósito**.
- Estos son los modelos utilizados por gobiernos y grandes instituciones para el análisis y la proyección de políticas públicas (pensiones, impuestos, salud).
- Todos estos modelos son descendientes intelectuales del trabajo original de **Guy Orcutt** (1957, 1961), quien propuso simular las unidades de decisión (familias) en lugar de agregados macro.

El Trabajo de Evaluación

Las presentaciones de la segunda evaluación se centrarán en analizar varios de estos modelos de MS dinámica en profundidad.

Modelo	País/Región	Propósito Principal (usos seleccionados)
DYNASIM 3	EE.UU.	Proyecciones de Seguridad Social, pensiones, análisis de riqueza.
CORSIM	EE.UU.	Acumulación de riqueza, salud (aunque ya no está activo).
DYNACAN	Canadá	Modelo de pensiones del gobierno, ya no está activo pero fue influyente.
LifePaths	Canadá	Marco genérico de Statistics Canada para simular historias de vida (salud, trabajo).
APPSIM	Australia	Proyecciones de pensiones, impuestos y beneficios.
MOSART	Noruega	Proyección de oferta laboral, educación y pensiones. Usado por Statistics Norway.
PENSIM2	Reino Unido	Proyecta el ingreso de los pensionados y evalúa reformas.
SESIM	Suecia	Proyecciones demográficas, pensiones y efectos distributivos.
SVERIGE	Suecia	Modelo espacial de población, demografía.
DESTINIE 2	Francia	Proyecciones de pensiones y análisis de reformas.
MIDAS	Bélgica / UE	Analiza la adecuación y sostenibilidad del sistema de seguridad social.
LIAM / LIAM2	Irlanda / UE	Marco genérico de Eurostat para comparar políticas de pensiones y transferencias.
POHEM	Canadá	Modelo de salud de Statistics Canada, evalúa costos y carga de enfermedad.

Fuente: Adaptado de Li, O'Donoghue y Dekkers (2014), Cap. 10.

Propósito Principal	Nro. de Modelos
Pensiones	34
Desigualdad y redistribución	13
Intergeneracional	6
Envejecimiento general	4
Demografía	10
Salud y cuidados a largo plazo	3
Educación	4
Espacial	5
Mercado laboral	1
Proyección de beneficios	1
Ahorro, riqueza y macro	5

Fuente: Li and O'Donoghue (2013).

MS Estática

- Evalúa el impacto **inmediato** de un cambio de política (ej. impuestos).
- **No** hay dimensión temporal.
- Los atributos de los agentes (edad, empleo) **no cambian**.
- Útil para análisis de "qué pasaría si..." en un año fiscal.

MS Dinámica

- Simula a los agentes a lo largo del **tiempo**.
- Los atributos de los agentes (edad, estado civil, empleo) **sí cambian**.
- Genera **trayectorias de vida** completas.
- Útil para analizar políticas a largo plazo (pensiones, educación).

Decisión 1: La Población Base

- La calidad del input determina la calidad del output. La elección de los datos iniciales es la primera gran decisión.
- **Opciones Comunes:**
 - **Datos de Encuestas:** Ricos en variables socioeconómicas (ej. ECH, EU-SILC), pero con muestras pequeñas y problemas de ponderación.
 - **Datos de Censos:** Gran cobertura y tamaño muestral, pero con menos variables de interés.
 - **Datos Administrativos:** Gran tamaño y precisión en variables clave (ej. ingresos fiscales), pero con información limitada en otras (ej. educación).
 - **Datos Sintéticos:** Se crean artificialmente para ajustarse a estadísticas conocidas. Útiles cuando no existen datos o para análisis teóricos.

Fuente de Datos	Porcentaje de Modelos (%)
Encuesta	43.8
Censo	23.4
Administrativa	17.2
Sintética	15.6

Fuente: Li and O'Donoghue (2013).

Decisión 2: Modelo de Cohorte vs. de Población

Modelo de Cohorte

- Simula una sola cohorte (ej. los nacidos en 1980) a lo largo de todo su ciclo de vida.
- **Pro:** Computacionalmente más simple. Permite un análisis detallado de una generación.
- **Contra:** No captura la interacción entre cohortes; menos útil para simular hogares e indicadores de pobreza.

Tendencia Actual

Con el aumento de la capacidad computacional, los modelos más nuevos tienden a ser modelos de población, ya que son más útiles para el análisis de políticas aplicadas.

Modelo de Población

- Simula una sección transversal de la población (todas las edades) a lo largo del tiempo.
- **Pro:** Permite la formación de hogares, interacciones entre individuos y el cálculo de indicadores agregados.
- **Contra:** Computacionalmente más intensivo.

Decisión 3: Envejecimiento Estático vs. Dinámico

"Envejecer" (aging) es el proceso de actualizar la base de datos para representar el futuro.

- **Envejecimiento Estático (Static Ageing):**

- **No** cambia los atributos de los individuos (ej. edad, estado civil).
- Solo ajusta los **ponderadores** de la muestra para que los agregados coincidan con proyecciones externas (ej. más peso a los viejos, menos a los jóvenes).
- **Desventaja:** No puede modelar la movilidad entre estados (ej. casarse, conseguir empleo) y por lo tanto no puede simular pensiones o historias de vida.

- **Envejecimiento Dinámico (Dynamic Ageing):**

- **Sí** cambia los atributos de los individuos a lo largo del tiempo.
- El modelo simula explícitamente eventos (muerte, nacimiento, empleo) para cada individuo y actualiza su estado.
- Es el enfoque estándar para la MS dinámica, ya que es el único que genera trayectorias de vida.

Decisión 4: Manejo del Tiempo (Discreto vs. Continuo)

Tiempo Discreto (El más común)

- El tiempo avanza en pasos fijos (ej. un año, un mes).
- En cada período, se simula la probabilidad de que ocurra un evento para cada individuo.
- **Desafío:** El **orden de los eventos** importa. ¿Simulamos primero el matrimonio o la fecundidad? (Una concepción puede afectar la probabilidad de conseguir empleo, y viceversa).
- **Limitación:** No maneja bien eventos de corta duración que ocurren *dentro* del período (ej. un desempleo de 3 meses en un modelo anual).

Tiempo Continuo (Eventos Discretos)

- El modelo no avanza año a año, sino de "evento a evento".
- Utiliza modelos de supervivencia (hazard) para simular el **tiempo exacto** hasta el próximo evento.
- **Desafío:** La estimación de los modelos de riesgos competitivos es muy exigente en términos de datos, y su calibración es más compleja.

Decisión 5: Modelos Abiertos vs. Cerrados

Esto se refiere a cómo el modelo maneja la creación de nuevas relaciones (ej. matrimonio).

- **Modelo Abierto (Open Model):**

- Si una agente (mujer) se casa, el modelo **crea un esposo sintético** "de la nada", asignándole atributos plausibles.
- **Pro:** Simple y rápido de computar (cada agente es independiente).
- **Contra:** La población puede dejar de ser representativa. Difícil de alinear a totales.

- **Modelo Cerrado (Closed Model):**

- Si una agente se casa, debe "encontrar" a su esposo **dentro de la población simulada** (ej. un hombre soltero de edad similar).
- **Pro:** Mantiene la consistencia de la población y es preferible para simular variables a nivel de hogar (ej. ingreso del hogar, pobreza).
- **Contra:** Mucho más complejo de implementar (requiere un "marriage market").

Decisión 6: ¿Cómo Modelar las Transiciones?

Tenemos un espectro de opciones para simular la $P(\text{evento}|X)$:

Enfoque Estadístico

- Modela las probabilidades directamente usando modelos estadísticos (ej. logits, hazards) o matrices de transición (cadenas de Markov).
- **Pro:** Relativamente simple de estimar e implementar si se tienen buenos datos de panel.
- **Contra:** Asume que las relaciones históricas son estables. No responde bien a cambios de políticas.

Enfoque Comportamental

- Modela el **proceso de decisión** del agente, usualmente basado en la teoría económica (ej. maximización de utilidad).
- Los parámetros de la política (ej. tasa de impuestos) entran directamente en la ecuación de decisión del agente.
- **Pro:** Necesario para evaluar el impacto de reformas de políticas, ya que captura la **respuesta comportamental**.
- **Contra:** Muy complejo de estimar y validar.

El Gran Desafío: Desviación (Drift) y Alineación

- **El Problema (Drift):** Los modelos de MS dinámica son complejos. Pequeños errores de especificación en las múltiples ecuaciones (mortalidad, fecundidad, empleo, etc.) se **acumulan** año tras año.
- Incluso un modelo bien especificado puede "desviarse" (drift) y, después de 20 años, producir resultados agregados (ej. tasa de empleo) que no tienen sentido o contradicen proyecciones macroeconómicas confiables.
- **La Solución Pragmática: Alineación (Alignment)**
 - La **Alineación** es el proceso de **restringir ex-post** los resultados del modelo para que coincidan con "blancos" (targets) agregados externos (ej. proyecciones oficiales de población o empleo).
 - Se ha vuelto una práctica estándar en casi todos los modelos de políticas a gran escala.

La alineación es, en esencia, un "parche" que corrige los resultados del modelo.

- **Variables Continuas (ej. Ingresos):**

- Generalmente se usa un escalado simple (multiplicativo o aditivo) para que la media simulada coincida con la media del target.

- **Variables Binarias (ej. Empleo):**

- Es más complejo. Si el modelo simula 1000 nuevos empleados pero el target dice que solo debería haber 900, ¿a quiénes "apagamos"?

- **Métodos comunes:**

- *Escalado Multiplicativo*: Se ajustan las probabilidades individuales antes del sorteo.
- *Ordenamiento (Sorting)*: Se ordenan los individuos por su probabilidad y se "encienden" los N con mayor probabilidad hasta alcanzar el target.

- **Riesgos de la Alineación:**

- Puede "torcer" las relaciones micro subyacentes de formas no deseadas.
- Puede forzar al modelo a alcanzar un target macro incluso si las condiciones micro lo hacen implausible.

Validación: ¿Cómo sabemos si el modelo es creíble?

La validación en MS es un proceso multi-etapa, y no existe un consenso único.

- **Validación de Componentes (Módulos):**

- ¿Los parámetros estimados en los módulos de transición (ej. logit de mortalidad) son estadísticamente significativos y tienen el signo esperado?

- **Validación Interna (Micro):**

- ¿Las distribuciones de las variables simuladas (ej. edad al primer hijo) se ven razonables? ¿Las correlaciones entre variables (ej. educación e ingresos) se preservan?

- **Validación Externa (Macro) o "Back-casting":**

- Esta es una prueba clave. Se inicializa el modelo en un año pasado (ej. 1970) y se corre "hacia adelante" hasta el presente (ej. 2020).
- Luego se compara la serie de tiempo simulada (ej. tasa de empleo) con la serie de tiempo histórica real.

- **Análisis de Sensibilidad:**

- ¿Qué tan sensibles son los resultados a cambios en los parámetros clave o en el orden de los eventos?

El capítulo 10 destaca varios desafíos que el campo aún enfrenta:

- **Transferencia de Conocimiento:** Históricamente, los modelos eran "cajas negras" en instituciones gubernamentales, con documentación pobre.
 - **Solución:** El movimiento hacia plataformas genéricas (ej. LIAM2) y *open-source* facilita la codificación y transferencia del conocimiento.
- **Altas Expectativas vs. Realidad:** Los modelos no son "bolas de cristal". La complejidad del comportamiento humano hace que las predicciones exactas sean imposibles.
 - **Solución:** Enfocarse en **análisis de escenarios** ("qué pasaría si...") en lugar de predicciones puntuales.
- **El "eslabón perdido" con ABM:** Los modelos de MS tradicionales son débiles para modelar interacciones sociales, redes y efectos de pares.
 - **Solución:** Integrar elementos de ABM para modelar estas interacciones explícitamente.

- En el laboratorio anterior, construimos un modelo "mecanicista" simple del proceso reproductivo.
- Este modelo es un híbrido entre un ABM (basado en reglas, generativo) y un modelo de MS (orientado a datos, busca ajustar tasas agregadas).
- En el próximo laboratorio exploraremos como generar "trayectorias" de mortalidad usando solo datos agregados. El objetivo es combinar ambas trayectorias (reproductiva / mortalidad) para generar historias de vida compatibles con tasas demográficas observadas. Evitando la necesidad de alinear.

Variable Aleatoria Discreta

Variable Aleatoria Discreta X

Valores potenciales: x_1, \dots, x_k

Función de Probabilidad:

$$f(x_k) = P(X = x_k) \quad k = 1, \dots, K$$

$$\sum_{k=1}^K f_k = \sum_{k=1}^K P(X = x_k) = 1$$

Función de Distribución:

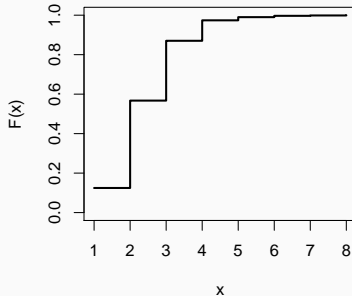
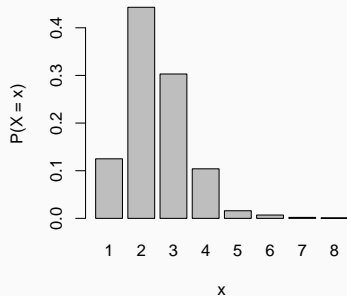
$$F(x) = P(X \leq x) = \sum_{x_k \leq x} f_k$$

Variable Aleatoria Discreta - Ejemplo

X: Nr. ideal de hijos

(IVS: Paises Europeos 1990)

x_k	1	2	3	4	5	6	7	8
f_k	0.125	0.443	0.303	0.104	0.016	0.007	0.002	0.001



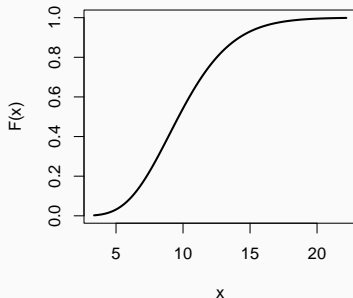
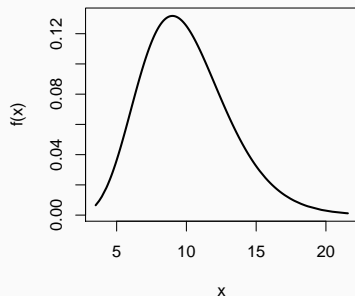
Variable Aleatoria Continua

Función de Densidad: $f(x)$

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

Función de Distribución:

$$F(x) = P(X \leq x)$$



Funciones para simular en R

`runif`: Distribución Uniforme
`rnorm`: Distribución Normal
`rgamma`: Distribución Gamma
`rexp`: Distribución exponencial

Ejemplos:

```
runif(n=100, min = 0, max = 1)  
rnorm(n=50, mean = 0, sd = 1)
```

Uno de los métodos para simular números aleatorios de una distribución:

Método de Inversión

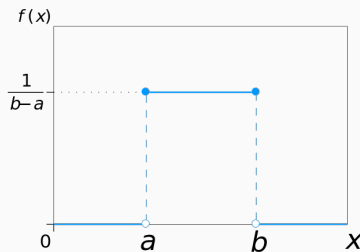


Figure 1: Distribución Uniforme

Teorema de la transformada inversa

Si $X \sim F(x)$
Entonces $U = F(X) \sim \mathcal{U}(0, 1)$
 \Downarrow

Si F es invertible y $U \sim \mathcal{U}(0, 1)$
Entonces $X = F^{-1}(U) \sim F(x)$

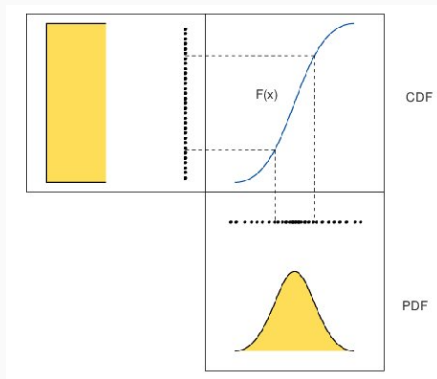


Figure 2: Método de Inversión

Método de Inversión

Para generar realizaciones de X :

1. Generamos valores aleatorios $u_1, \dots, u_n \sim U(0, 1)$.
2. Aplicamos la transformación: $x_i = F^{-1}(u_i)$ para $i = 1, \dots, n$.

De esta manera, los x_i obtenidos siguen la distribución de probabilidad deseada.

Ejemplo:

Supongamos que queremos generar una variable aleatoria exponencial con parámetro λ :

- La función de distribución acumulada es $F(x) = 1 - e^{-\lambda x}$
- Su inversa es $F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u)$

Entonces:

1. Generamos $u \sim U(0, 1)$.
2. Calculamos $x = F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u)$.

Generación de números aleatorios de la distribución exponencial a partir del método de la transformación inversa.