Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática

Laboratorio 1

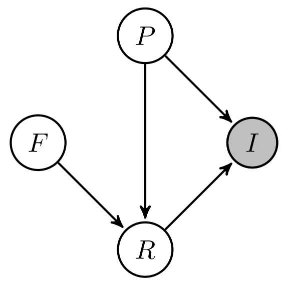
Estudiante: Daniel Minaya
Profesor: Felipe Tobar
Auxiliares: Cristóbal Alcázar
Camilo Carvajal

Fecha de entrega: 8 de septiembre

Parte 2 (Modelos gráficos)

(a) Sabemos que la cantidad de polen puede causar irritación en nariz y ojos en algunos casos, por lo que habría una dependencia entre estas dos variables. Además, la cantidad de polen también afecta a la probabilidad de sufrir Rinitis alérgica, por lo que tendríamos otra dependencia. También es sabido que si se tiene Rinitis alérgica, entonces es probable tener irritación en nariz y ojos. Por último, suponiendo que la variable de tener un familiar con Rinitis alérgica hace referencia a una enfermedad hereditaria, y no a una enfermedad que pueda ser contagiada, tendríamos que esta variable aumentaría las posibilidades de tener la Rinitas alérgica.

Definiendo las variables aleatorias $I =$ "Tener irritación en nariz y ojos", $P =$ "cantidad de polen en el ambiente", $R =$ "Tener Rinitis alérgica" y $F =$ "Tener un familiar con Rinitis alérgica", entonces, por todo lo anterior, el modelo gráfico que representa el problema estaría dado por el siguiente DAG:



(b) Usando la notación anterior, queremos encontrar $\mathbb{P}(R|I)$. Según el modelo gráfico planteado, la probabilidad conjunta está dada por:

$$\mathbb{P}(P, I, F, R) = \mathbb{P}(P) \cdot \mathbb{P}(F) \cdot \mathbb{P}(R|F, P) \cdot \mathbb{P}(I|R, P),$$

donde todas las probabilidades en el lado izquierdo son conocidas a priori. De este modo, por el teorema de Bayes tenemos:

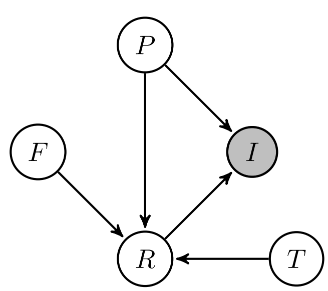
$$\mathbb{P}(R|I) = \frac{\mathbb{P}(I|R)\mathbb{P}(R)}{\mathbb{P}(I)},$$

luego, por el teorema de probabilidades tenemos:

$$\begin{aligned}
\mathbb{P}(I|R) &= \sum_k \mathbb{P}(I|R, P = k)\mathbb{P}(P = k), \\
\mathbb{P}(R) &= \sum_k \sum_j \mathbb{P}(R|F = j, P = k)\mathbb{P}(F = j)\mathbb{P}(P = k), \\
\mathbb{P}(I) &= \sum_k \sum_j \mathbb{P}(I|R = j, P = k)\mathbb{P}(R = j)\mathbb{P}(P = k),
\end{aligned}$$

donde las primeras dos igualdades están en términos de probabilidades conocidas, mientras que la última igualdad tiene el término $\mathbb{P}(R = j)$ que se conoce de la segunda igualdad, por lo que $\mathbb{P}(R|I)$ se obtendría calculando las probabilidades respectivas.

(c) En este caso debemos agregar una nueva variable al modelo gráfico, llamemósla $T =$ "Test positivo". Esta variable solo afecta a la probabilidad de tener Rinitis alérgica, por lo que el nuevo modelo gráfico estaría dado por:



Con este nuevo modelo gráfico, la nueva probabilidad conjunto sería:

$$\mathbb{P}(P, I, F, R, T) = \mathbb{P}(P) \cdot \mathbb{P}(F) \cdot \mathbb{P}(T) \cdot \mathbb{P}(R|F, P, T)\mathbb{P}(I|R, P),$$

luego la probabilidad $\mathbb{P}(R|I) = \frac{\mathbb{P}(I|R)\mathbb{P}(R)}{\mathbb{P}(I)}$ solo vería afectada su término $\mathbb{P}(R)$, el cual ahora estaría dado por

$$\mathbb{P}(R) = \sum_k \sum_j \sum_i \mathbb{P}(R|T = i, F = j, P = k)\mathbb{P}(T = i)\mathbb{P}(F = j)\mathbb{P}(P = k).$$

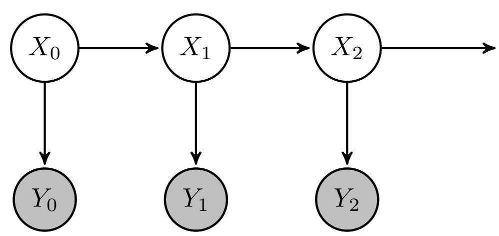
Notar que esto afectaría indirectamente a $\mathbb{P}(I)$ también, pues depende de $\mathbb{P}(R)$.

(d) En este caso utilizamos P como variable continua que sigue una distribución semi-normal, pues al tratarse de la cantidad de polen en el ambiente, ésta puede tomar solo valores positivos. En este caso, la expresión para $\mathbb{P}(R|I)$ cambia, pues pasamos P de una variable discreta a una continua, por lo que:

$$\begin{aligned}
\mathbb{P}(I|R) &= \int_0^\infty \mathbb{P}(I|R, P = x)f_P(x)dx, \\
\mathbb{P}(R) &= \int_0^\infty \sum_j \sum_i \mathbb{P}(R|T = i, F = j, P = x)\mathbb{P}(T = i)\mathbb{P}(F = j)f_P(x)dx, \\
\mathbb{P}(I) &= \int_0^\infty \sum_j \mathbb{P}(I|R = j, P = x)\mathbb{P}(R = j)f_P(x)dx,
\end{aligned}$$

donde $f_P(x)$ es la densidad de una semi-normal, es decir, $f_P(x) = \sqrt{\frac{2}{\sigma^2 \pi}} e^{-x^2/2\sigma^2}$.

(e) Definamos X_t = "cantidad de polen en el tiempo t" e Y_t = "irritación en nariz y ojos en el tiempo t" procesos aleatorios. Como suponemos que el futuro es independiente del paso condicionado al presente, entonces $(X_t)_t$ es una cadena de Markov, por lo que nuestro modelo gráfico ahora es un Hidden Markov Model representado por:



Queremos predecir la cantidad de polen cada 30 minutos usando como única medición el nivel de irritación en nariz y ojos, es decir, se quiere calcular $\mathbb{P}(X_{30t}|Y_{1:30t}) = \mathbb{P}(X_{30t}|Y_1, \ldots, Y_{30t})$, para lo cual se podría utilizar el filtro de Kalman.