Inferencia Bayesiana en la distribución Gumbel: aplicación en el modelamiento de intensidades de lluvia

Ignacio Vidal G.

.

ividal@utalca.cl

Universidad de Talca, Chile.

Seminario FONDEF-D08I1054, Santiago de Chile.
8 de enero de 2013

Esquema

Esquema

Esquema

Esquema

Es importante conocer la frecuencia de intensidades máximas de lluvia

La intensidad de precipitación representa la cantidad de lluvia caída en función del tiempo, expresada comúnmente en milímetros por hora (mm/hr).

Las intensidades de precipitación muy altas son poco frecuentes, pero cuando ocurren pueden provocar grandes desastres tales como inundaciones, desbordes de ríos, deslizamiento de laderas, etc.

Estos problemas se pueden mitigar o evitar con un buen diseño de obras hidráulicas, los que necesariamente requieren del conocimiento de las intensidades máximas de precipitación.

En hidrología se construyen las curvas de IDF

Las intensidades de lluvias están estrechamente relacionadas con el tiempo de duración de las lluvias: las mayores intensidades están asociadas a duraciones menores.

Además, también es importante conocer la frecuencia con que ocurren las mayores intensidades de lluvia para una duración específica.
Por esta razón es que cobra importancia el estudio de la relación entre estas tres variables: intensidad, duración y frecuencia.
De aquí surge la importancia de contar con las curvas de intensidad-duración-frecuencia (curvas IDF).

Una vez determinados los patrones de intensidad de lluvia en función de los períodos de retorno (o frecuencias) y la duración, se pueden diseñar mapas de inundaciones y obras hidráulicas.

Debíamos construir las curvas IDF en la zona centro sur de Chile

Este trabajo surge por la necesidad de contar con las curvas IDF en la zona centro sur de Chile y contar con un sistema de actualización semi-automático y permanente en el tiempo de estas curvas.

Como la distribución Gumbel es la que se ha utilizado tradicionalmente para ajustar las intensidades máximas de lluvia en Chile, nosotros también debíamos utilizar esta distribución en nuestro trabajo.

Los pocos datos apoyan la aplicación de la metodología Bayesiana

El otro aspecto interesante de este trabajo es que hay muchas estaciones pluviográficas que cuentan con pocos años de registro de datos.

En este sentido el análisis Bayesiano nos da una gran ayuda porque utilizamos la información de estaciones pluviográficas similares, como información a priori para estimar las frecuencias de una estación en particular dentro del grupo de estaciones pluviográficas similares.

Un experto definió cuáles eran los grupos de estaciones similares de acuerdo a diferentes variables tales como altitud, latitud, humedad, etc.

Recolección de datos en cada estación pluviográfica

Si $I(t)$ es la intensidad de lluvia en el tiempo t, entonces se observa $Y_k(d) = \int_k^{k+d} I(t)dt$ que es la lluvia caída en el intervalo de tiempo $(k, k+d)$.
Por ejemplo, si $d = 60$ min, entonces para cada día $j = 1, \ldots, 365$ del año i se observa $Y_{ij1}(60), Y_{ij2}(60), \ldots, Y_{ij24}(60)$. Luego, para el i-ésimo año se reporta $M_i(d) = \max_{j,k} \{Y_{ijk}(d)\}$ que sería el máximo de lluvia anual para una duración d.
Sin embargo, lo común en la práctica es trabajar con las intensidades máximas medias por hora, o sea

$$X_i(d) = \frac{M_i(d)}{d/60}, \qquad \text{mm/hr.}$$

Como el análisis de frecuencia que haremos en este trabajo es el mismo para todo d, entonces de ahora en adelante suprimiremos la letra d.

Con estos supuestos es factible considerar el modelo Gumbel

En este trabajo vamos suponer que para cada (i,d) fijo, $Y_{ijk}(d)$ son variables aleatorias *iid* de una distribución cuyas colas tienen una caída exponencial.

Además, supondremos que $X_1,\ldots,X_n$ es una muestra aleatoria de las intensidades máximas anuales de lluvia en una determinada estación pluviográfica.

Por tanto supondremos que $X_1,\ldots,X_n \overset{iid}{\sim} Gumbel(\mu,\sigma)$.
La función de densidad de probabilidad (pdf) es

$$f(x\,|\mu,\sigma) = \frac{1}{\sigma}e^{-(x-\mu)/\sigma}\exp\left[-e^{-(x-\mu)/\sigma}\right],$$

donde $x\in\mathbb{R}$, $\mu\in\mathbb{R}$ y $\sigma\in\mathbb{R}_+^*$.

Así seleccionamos las distribuciones a priori

Con el objetivo de hacer inferencias a posteriori sin la necesidad de contar permanentemente con un experto en el tema y a la vez involucrar información de varias estaciones pluviográficas similares, es que se condieraron las siguientes distribuciones a priori.

- La distribución no informativa $f(\mu|\sigma) \propto \sigma^{-1}$ para el parámetro μ
- $f(\sigma) \propto \sigma^{-a-1} e^{-b/\sigma}$, donde $a > 0$ y $b > 0$, o sea $\sigma \sim IG(a, b)$.

Para cada estación pluviográfica, los hiperparámetros a y b fueron tomados de forma tal que $\mathbb{E}(\sigma) = s\sqrt{6}/\pi$ y $\mathrm{var}(\sigma) = Cs\sqrt{6}/\pi$, donde $C \in \mathbb{R}_{+}^{*}$ y s es la desviación estándar muestral.
En este trabajo s se calculó utilizando los datos de todas las estaciones de medición consideradas homogéneas (o similares) según el criterio de un experto.

Todas nuestras inferencias se basan en la distribución a posteriori

Habiendo observado la muestra aleatoria $\mathbf{x} = (x_1, \ldots, x_n)$ de una distribución *Gumbel* (μ, σ) y para las distribuciones a priori establecidas anteriormente, la distribución a posteriori de (μ, σ) es

$$f(\mu, \sigma \,|\mathbf{x}) = A \frac{1}{\sigma^{n+a+2}} \exp\left(-\frac{n(\bar{x} - \mu) + b}{\sigma} - e^{\mu/\sigma} \sum_{i=1}^{n} e^{-x_i/\sigma} \right),$$

donde A es la constante de normalización.

Todas las medidas de resumen a posteriori pueden aproximarse por el método de Laplace

A partir de la posteriori pudimos calcular:

- La constante de normalización A.
- Las modas a posteriori de (μ, σ).
- Medias, varianzas, covarianzas e intervalos de credibilidad a posteriori de (μ, σ).
- La pdf de la distribución predictiva $f(x_{n+1} | \mathbf{x})$, su valor esperado, su distribución acumulada $F(x_{n+1} | \mathbf{x})$ y sus intervalos de credibilidad.

Un experto determinó las estaciones pluviográficas similares a la de Talca

Figura: 18 estaciones pluviográficas en la región del Maule en Chile.

Para cantidades de lluvia tomadas en intervalos de tiempo de una hora,

las distribuciones a priori fueron:

- La distribución no informativa $f(\mu|\sigma) \propto \sigma^{-1}$ para el parámetro μ
- $\sigma \sim IG(a, b)$.

Los hiperparámetros a y b fueron tomados de forma tal que $\mathbb{E}(\sigma) = s\sqrt{6}/\pi$ y $\mathrm{var}(\sigma) = (100)\, s\sqrt{6}/\pi$.

Tomamos el valor $s = 3{,}99669 \approx 4$ que fue la desviación estándar de las intensidades máximas de lluvia anuales de todas las estaciones pluviográficas similares a la de Talca.

Registro de 28 años comprendidos desde 1982 hasta 2009

Estimaciones	μ	σ
mle	11,02	2,77
Intervalo de confianza	(9,94; 12,10)	(2,07; 3,70)
Media a posteriori	10,96	2,79
Moda a posteriori	10,95	2,61
Intervalo de credibilidad	(9,87; 12,11)	(1,93; 3,40)

Cuadro: Estimación de parametros para intensidades máximas de lluvia anuales en Talca, Chile.

Registro de 28 años comprendidos desde 1982 hasta 2009

Estimaciones	μ	σ
mle	11,02	2,77
Intervalo de confianza	(9,94; 12,10)	(2,07; 3,70)
Media a posteriori	10,96	2,79
Moda a posteriori	10,95	2,61
Intervalo de credibilidad	(9,87; 12,11)	(1,93; 3,40)

Cuadro: Estimación de parametros para intensidades máximas de lluvia anuales en Talca, Chile.

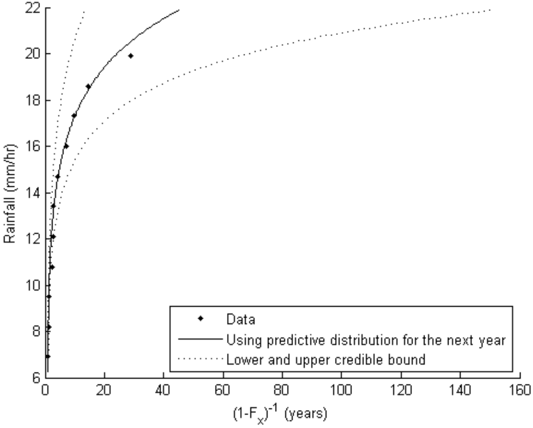
Por otro lado, la estimación del valor máximo de intensidad de lluvia esperado para el próximo año fue

$$\mathbb{E}\left(\widehat{x_{n+1}}\,|\mathbf{x}\right) = 12{,}57 \qquad \text{mm/hr.}$$

Con un intervalo de credibilidad al 95 % igual a

$$(7{,}17; 21{,}52)$$

Período de retorno: $T(x) = 1/\mathbb{P}(X > x)$



Por ejemplo, en un período de 10 años se epera ocurra al menos un evento de intensidad máxima anual igual a 17,32 mm/hr.

Las curvas IDF son casi iguales para duraciones menores a 45 minutos

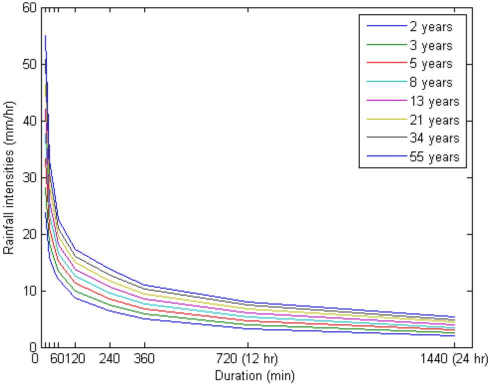


Figura: Curvas IDF empíricas basadas en intensidades máximas de lluvia anuales en Talca, Chile.

La solución a este problema tenía tres restricciones prácticas:

1. Seguir usando la distribución de Gumbel.
2. Ofrecer un método de estimación permanente en el tiempo que fuera capaz de acutalizarse en forma automática sin la presencia necesaria de un experto.
3. Estaciones pluviográficas con poca cantidad de registro de datos.

Nuestra solución tiene en cuenta las tres restricciones anteriores

Hemos asignado distribuciones a priori para los parámetros de la distribución de Gumbel de forma tal que:

1. Son capaces de utilizar la información de un experto para ayudar a la estimación en localidades con poca cantidad de datos.

2. Como son poco informativas, sirven como distribuciones a priori implícitas para mantener el proceso de estimación actualizándose permanente en el tiempo.

3. Para obtener las estimaciones a posteriori no es necesario el uso de métodos de simulaciones ni la presencia permanente de un estadístico Bayesiano.

Los tomadores de desiciones tienen la posibilidad de utilizar la metodología Bayesiana o la de los mle

Nosotros recomendamos utilizar la metodología Bayesiana salvo que se quieran obtener resultados específicos para una estación pluviográfica sin que los datos de otras estaciones similares influyan en dichos resultados.

Otros aspectos que debieran abordarse en trabajos posteriores son:

- Utilizar distribuciones más generales para la modelación de la frecuencia de eventos extremos, como por ejemplo, la distribución de valores extremos generalizada.
- Estimar directamente las curvas IDF y no utilizar el método tradicional.
- En el proceso de estimación de las curvas IDF las variables aleatorias $X_i\,(d_1)$ y $X_i\,(d_2)$, con $d_1 \neq d_2$, para cada año i no son independientes. Aunque eso no afecta el trabajo presentado en este artículo, sí influye a la hora de estimar las curvas IDF.
- Otra vía para estimar las curvas IDF puede ser buscar directamente la distribución a posteriori de los períodos de retorno.