

Reporte práctica 4

Diagramas de Voronoi

28 de agosto de 2017

1. INTRODUCCIÓN

Esta práctica aborda la simulación del crecimiento de grietas sobre un material. El material se encuentra dividido en regiones conocidas como celdas de Voronoi que se asemejan a núcleos formados por un proceso de cristalización del material. Si representamos el material con una cuadrícula bidimensional, una celda de Voronoi está formada por las casillas más próximas a cada una de las k semillas inicialmente dadas. Una grieta comienza en las orillas de la cuadrícula y tiene mayor probabilidad de propagarse por las fronteras de las celdas de Voronoi que por el interior de las mismas. También se supone que la grieta se propaga por la frontera siempre que pueda hacerlo. Se desea estudiar el efecto que tiene el número de semillas y el tamaño de la cuadrícula sobre los tamaños de las grietas que se forman.

2. EXPERIMENTO

El experimento se desarrolló variando los tamaños de la cuadrícula $n \in \{50, 100, 500, 1000, 2000, 4000\}$. La cantidad de semillas k se considera proporcional al tamaño de la cuadrícula, para poder visualizar un efecto que no dependa de la escala. Se considera $k = \alpha n$, donde $\alpha \in \{\frac{1}{2}, 1, 2, 4, 8, 16, 32\}$ y se realizaron 200 réplicas. La Figura 3.1 muestra la distribución de los largos alcanzados por las grietas. Para una mejor visualización, se graficó en el eje del largo de las grietas solo hasta 200. La línea roja representa la media de los largos para cada tamaño fijo, esto para complementar la falta de datos sin visualizar, sobre todo en la celda correspondiente a una cuadrícula de 50×50 . En el caso de las cuadrículas de tamaño 4000, sólo se muestra los largos de grietas obtenidas con 2000, 4000 y 8000 semillas.

3. RESULTADOS

Observe como la media de los largos disminuye a medida que lo hace el tamaño de la cuadrícula. La razón es muy sencilla de explicar, a medida que el tamaño de la cuadrícula aumenta, lo hacen también el tamaño de las celdas de Voronoi. Como la probabilidad de que una celda cualquiera esté dentro de la celda y no en la frontera, y la grieta no se expande fácilmente sobre el interior, es de esperar que sean muy cortas. Una prueba Kruskal y Wallis con significancia de 0.001 muestra que sí hay diferencia estadística de los largos de las grietas respecto a los tamaños de las cuadrículas. Además, una prueba de Dunn con la misma significancia, indica que cada par de niveles es diferente estadísticamente.

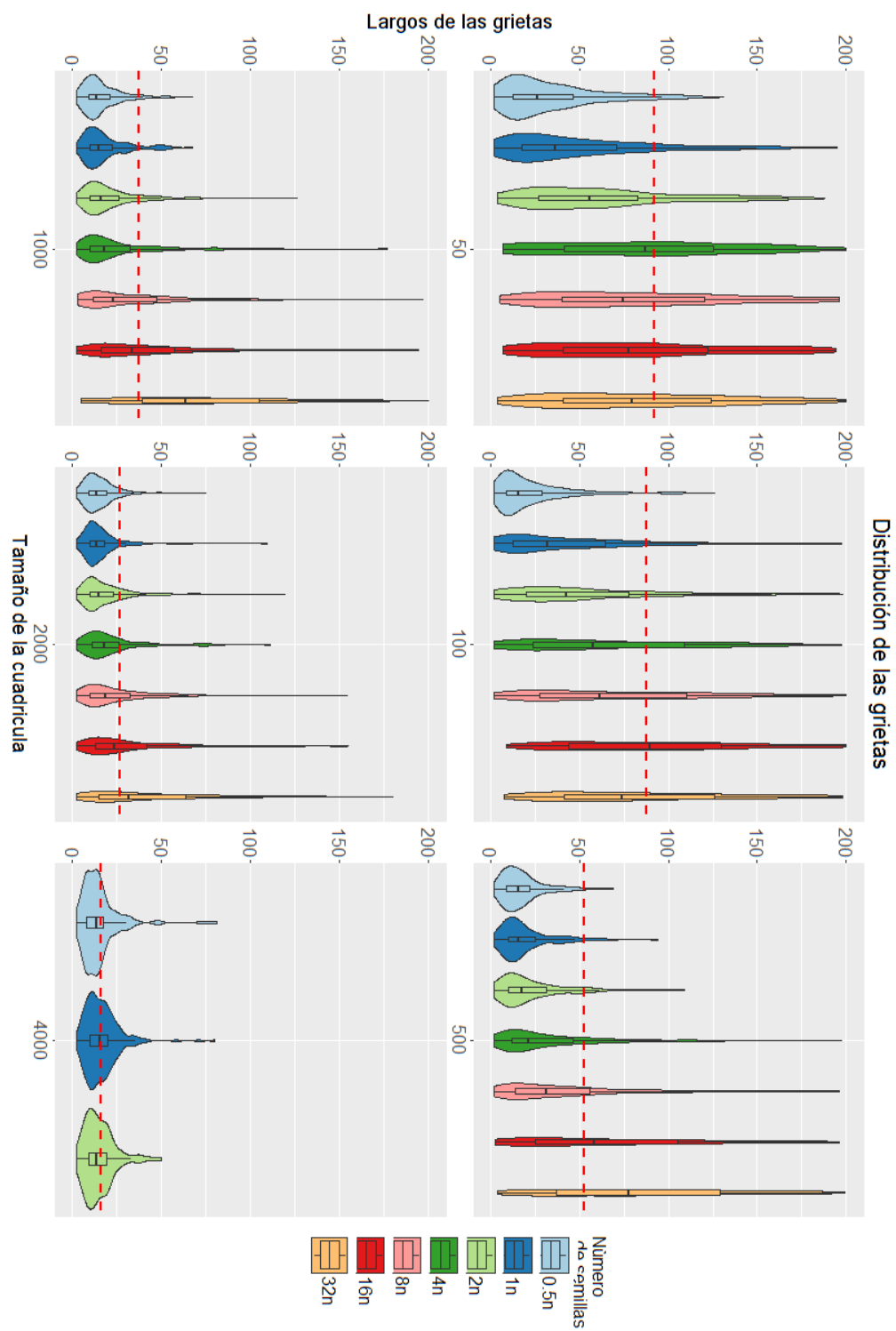


Figura 3.1: Diagramas de violín y de bigotes correspondientes a los largos de las grietas respecto al tamaño de la cuadrícula y la cantidad inicial de semillas.

Respecto al número de semillas, se espera que sin importar el tamaño de la cuadrícula el comportamiento de los resultados sea similar, esto por la proporcionalidad. Podemos ver que a medida que el número de semillas aumenta, también lo hacen las dimensiones de las grietas. La razón es la misma que cuando analizamos los tamaños de las cuadrículas; si hay más semillas, por ende más núcleos y por tanto más celdas que son frontera. Es decir; hay mayor probabilidad de que las grietas aumente. Podemos ver también este fenómeno reflejado en la forma de la distribución (violines). Cuando hay pocas semillas, se espera que las grietas sean pequeñas, pero hay mayor variabilidad dependiendo de si tocó una frontera o no. Cuando hay muchas semillas y es más probable encontrar una a frontera, el crecimiento es más estable.

4. RETO 1. VARIANDO LAS DISTRIBUCIONES DE LAS SEMILLAS

Ahora vamos a analizar el efecto que tiene la forma en que se distribuyen las semillas en un inicio. Como se hacía anteriormente, las semillas se distribuían uniformemente sobre toda la cuadrícula, ahora proponemos cinco nuevas distribuciones iniciales. Para lograrlo, hacemos combinaciones de la probabilidades con que la función `sample(x, size, replace = FALSE, prob = NULL)` permite obtener una fila o columna con una distribución introducida en `prob`. Para definir los vectores de probabilidades se calcularon los n valores $F_{i+1} - F_i$, donde F representa una distribución acumulada e $i \in \{1, \dots, n\}$. Las distribuciones son las siguientes:

Tina de baño: Esta distribución de probabilidades se asemeja en forma a una tina para bañarse. La idea es que la probabilidad de seleccionar, por ejemplo una fila, sea mayor para las primeras y/o últimas filas. Esta distribución se logró con las siguientes instrucciones:

```
#vector de probabilidades tipo bañera
t=pexp(q=seq(0,10,10/(n/2)))
ex=sapply(2:length(t),function(x) t[x]-t[x-1])
ex=c(ex,rev(ex))/2
```

La Figura 4.1a ilustra la distribución.

Normal: La clásica distribución normal nos permite seleccionar con mayor probabilidad las filas y/o columnas del centro de la cuadrícula. Específicamente, se utiliza una distribución con media $\mu = 0$ y varianza $\sigma = 0.5$. La Figura 4.1b ilustra la distribución generada.

```
#vector de probabilidades normal
t=pnorm(q=seq(-3,3,6/n),sd=0.5)
nor=sapply(2:length(t),function(x) t[x]-t[x-1])
```

De igual forma se obtuvo una distribución para seleccionar únicamente sobre la primera mitad de las filas y/o columnas. Se adjuntó la homóloga para obtener la segunda mitad y se obtuvo una distribución como la que aparece en la Figura 4.1c. El nombre de este vector de probabilidades es `nor2`.

Gamma: Una distribución gamma se utiliza para seleccionar las filas y columnas en un grupo descentrado. La función de R que genera la acumulada de esta distribución permite

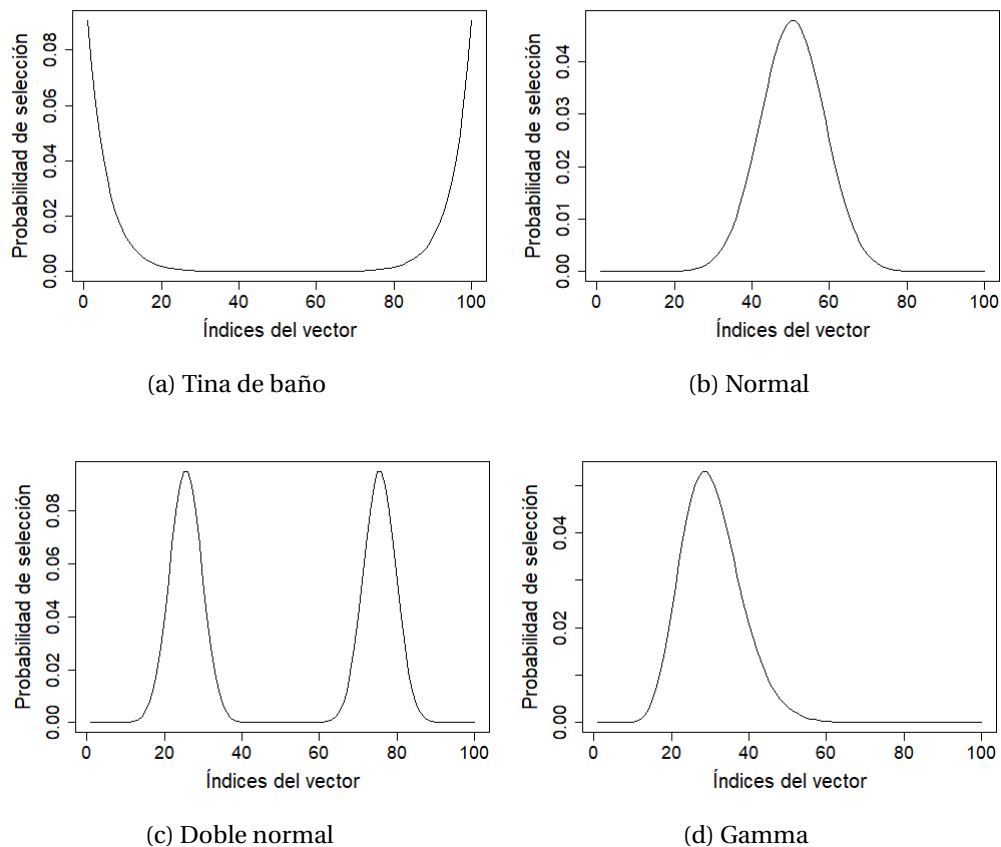


Figura 4.1: Ejemplos de distribuciones utilizadas como vectores de probabilidad para el posicionamiento de las semillas en una cuadrícula de 100×100 .

configurar la forma (shape) y la manera de escalarla (rate). La Figura 4.1d muestra una distribución obtenida con `shape=15` y `rate=5`.

```
#vector de probabilidades tipo gamma
t=pgamma(q=seq(0,10,10/n),shape=15,rate=5)
gam=sapply(2:length(t),function(x) t[x]-t[x-1])
```

Con estos vectores de probabilidades, se hacen combinaciones para lograr mayor probabilidad de selección en ciertas zonas de la cuadrícula. Estas son las cinco nuevas distribuciones para las semillas:

Orillas: Las semillas se distribuyen con mayor probabilidad sobre las orillas de la cuadrícula. Esto se logra seleccionando una fila de acuerdo a la tina de baño y su correspondiente columna uniformemente. Esto por supuesto, sólo nos daría las orilla de arriba y abajo de la cuadrícula. Sin embargo, con la misma probabilidad se puede seleccionar una

columna exponencialmente y su fila correspondiente de manera uniforme, obteniendo las cuatro orillas. Las instrucciones son las siguientes:

```
if(tipo==1){ #las semillas se distribuyen sobre la orilla
  if(runif(1)<0.5){
    fila <- sample(1:n, 1,prob=ex)
    columna <- sample(1:n, 1)
  }else{
    fila <- sample(1:n, 1)
    columna <- sample(1:n, 1,prob=ex)
  }
  nombre="orillas"
}
```

La Figura 4.2a muestra la distribución de las semillas y el diagrama de Voronoi correspondiente.

Centro: Seleccionando de acuerdo a una distribución normal las filas y columnas de la cuadrícula, las posiciones mas probables para las semillas serán en el centro de ésta. La Figura 4.2b muestra un ejemplo de la distribución de semillas y el diagrama de Voronoi correspondiente.

Esquinas: La distribución tina de baño, nos permite seleccionar con mayor probabilidad las esquinas de las cuadrículas como inicio para las semillas; formando núcleos como los de la Figura 4.2c.

Grupos: Haciendo uso de la distribución que llamamos doble normal y de la simetría de la cuadrícula, se pueden formar cuatro grupos de celdas de mayor probabilidad de selección para el posicionamiento de las semillas. Ésta idea se podría replicar y formar fácilmente $2g, g \in \mathbb{N}$ grupos. Sin embargo; cuando g se incrementa, la distribución tiende a ser uniforme. Por motivos de comparación, sólo se consideró este caso. La figura 4.2d muestra la posición inicial de las semillas y el diagrama de Voronoi correspondiente.

Grupo descentrado: La idea de formar un grupo que no estuviera en el centro era deformar las celdas de Voronoi que tocan la frontera de la cuadrícula. Esto se logra seleccionando filas y columnas con la distribución tipo Gamma. La Figura 4.2e muestra un ejemplo de diagrama Voronoi obtenido.

Uniforme: Este es el caso estándar en donde las filas y columnas pueden seleccionarse de igual forma en cualquier posición de la cuadrícula. Se muestra el diagrama de Voronoi que se puede obtener con esta distribución en la Figura 4.2f.

4.1. EXPERIMENTO Y RESULTADOS

Para determinar si la distribución inicial de las semillas tiene efecto sobre los tamaños alcanzados por las grietas, realizamos un experimento donde variamos las seis distribuciones iniciales ya mencionadas (véase Figura 4.2) Se consideran cuadrículas de largo 50, 100 y 500

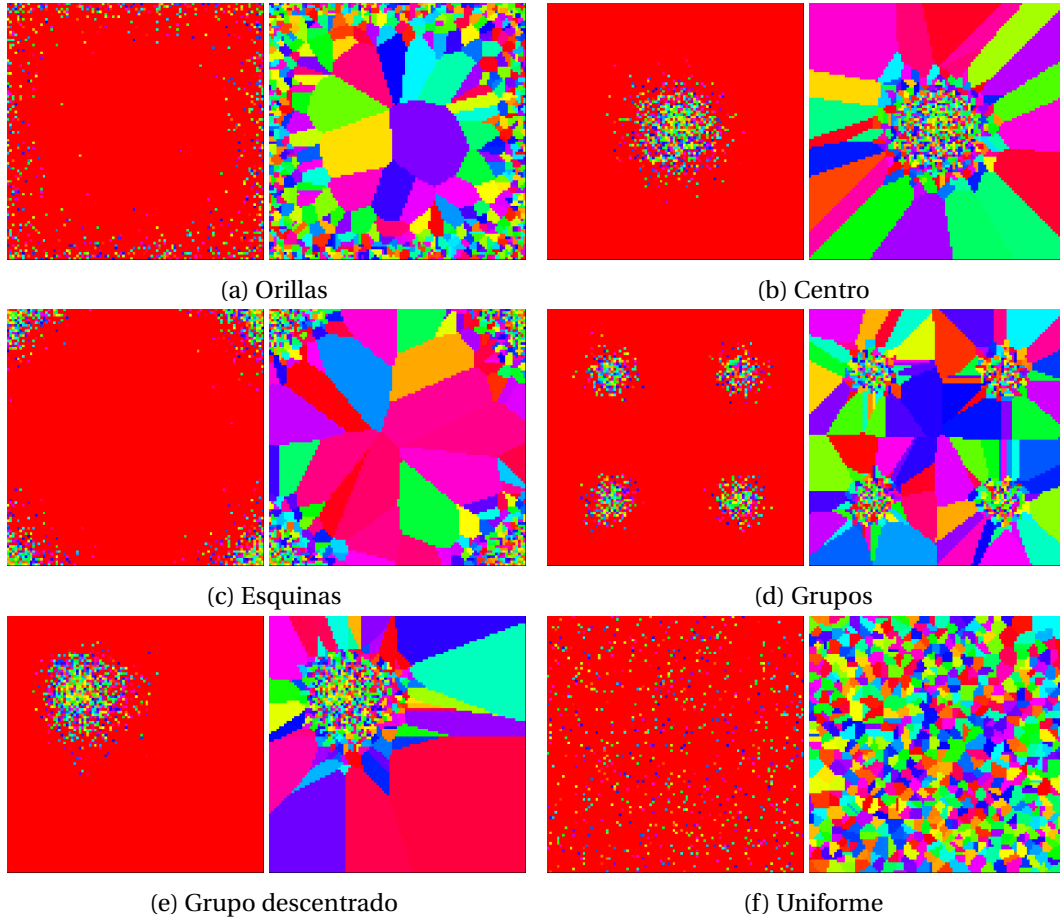


Figura 4.2: Distribución inicial de semillas y diagrama de Voronoi correspondiente para cada una de los cinco vectores de probabilidades utilizados en una cuadrícula de 100×100 y con 800 semillas.

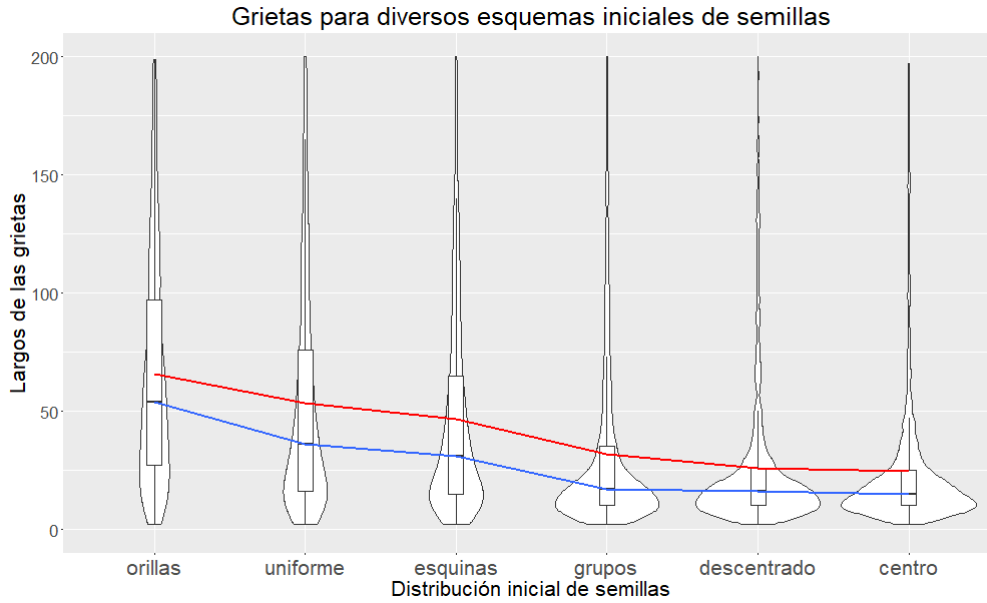


Figura 4.3: Distribución de los largos de las grietas formadas cuando se varia la forma en que las semillas son colocadas inicialmente.

y una cantidad de semillas proporcional al tamaño de la cuadrícula, se consideran $k = pn$ semillas, donde $p \in \{0.5, 1, 2, 4, 8, 16, 32\}$. El comportamiento del tamaño alcanzado por las grietas en dependencia de la configuración inicial de las semilla se aprecia en la Figura 4.3. Anteriormente, ya se había probado la significancia tanto del tamaño como del número de semillas para el tamaño de las grietas. La variación de éstos dos últimos factores se hizo para capturar mayor variabilidad en los resultados.

Una prueba de Kruskal y Wallis con un valor p de 2.2×10^{-16} indica la significancia de la forma en que se distribuyen las semillas para los largos de las grietas. Podemos notar como hay cierta similitud entre las configuraciones donde hay semillas en el exterior (*orillas*, *esquinas* y *uniforme*) y entre aquellas donde las semillas están dentro de la cuadrícula (*centro*, *grupos* y *descentrado*). La razón es la forma en que inician las grietas, siempre en una orilla de la cuadrícula, y la facilidad de crecer por la frontera de las celdas y no por su interior. Se puede esperar que entre más probable sea que la grieta encuentre una frontera y más en la orilla, ésta podría propagarse más.

Para comprobar si hay configuraciones estadísticamente equivalentes, se hace una prueba de Dunn, donde con un valor p de 0.44 se indica la equivalencia estadística del nivel *centro* y el nivel *descentrado*. Todos los demás configuraciones son distintas. La línea azul en la Figura 4.3 une a las medianas de cada configuración mostrando la tendencia. La línea roja muestra unidas las medias de cada configuración, esto porque los límites de la gráfica en el eje vertical fueron cortados a 200 para una mejor visualización. Puede observarse que colocar las semillas en las orillas hace que las grietas sean más grandes, mientras que hacer un sólo grupo de semillas en el interior de la cuadrícula afecta en gran medida el largo de las semillas, teniendo sólo como datos atípicos las grietas de gran tamaño.

5. RETO 2: AGREGANDO SEMILLAS PAULATINAMENTE

La siguiente variante es analizar el tamaño de las grietas en diagramas de Voronoi donde los núcleos se formaron paulatinamente alrededor de las semillas y éstas se fueron incluyendo una a una. La tasa de crecimiento de los núcleos se considera constante mientras que la probabilidad de incluir una semilla es un nuevo factor del experimento. En una iteración del proceso de crecimiento de los núcleos, una celda no asignada se asigna a un núcleo i si y sólo si se encuentra dentro de un radio r_i de la semilla que formó al núcleo.

```
cellda <- function(pos) {  
  fila <- floor((pos - 1) / n) + 1  
  columna <- ((pos - 1) %% n) + 1  
  if (voronoi[fila, columna] > 0) { # ya tiene una semilla  
    return(voronoi[fila, columna])  
  } else {  
    for (semilla in sample(1:length(radio))) {  
      dx <- columna - x[semilla]  
      dy <- fila - y[semilla]  
      dist <- sqrt(dx^2 + dy^2)  
      if (dist <= radio[semilla]) {  
        return(semilla)  
      }  
    }  
  }  
  return(voronoi[fila, columna])  
}
```

Los radios r_i aumentan con una tasa de $0.15n$, donde n es el largo de la cuadrícula. Una nueva semilla se puede incluir con probabilidad p_n y puede colocarse en una celda que no este asignada. La posición propuesta de colocación se distribuye de acuerdo a las configuraciones presentadas en la sección 4, exceptuando la llamada *grupo descentrado*, pues estadísticamente lo representa la configuración *centro*. Cuando todas las celdas sean asignadas a un núcleo completamos el diagrama de Voronoi. Un ejemplo del crecimiento de estas celdas de Voronoi aparece en la Figura 5.1 donde las semillas aparecen con mayor probabilidad en el centro de la cuadrícula. El tamaño de ésta última es de 500×500 y las semillas aparecen con probabilidad del 60% en cada iteración.

Adjunto a este reporte, se encuentran archivos gif que ilustran el crecimiento para los distintos tipos de distribuciones de las semillas.

5.1. EXPERIMENTO Y RESULTADOS

Se realizó un experimento para comprobar la significancia de la cuadrícula, la probabilidad de agregar una nueva semilla y el tipo de distribución para colocar la semilla sobre el largo de las grietas formadas sobre el diagrama de Voronoi. Se consideraron cuadrículas de 250, 500 y 1000 celdas de largo, probabilidades de una nueva semilla de 0.2, 0.4, 0.6 y los tipos de distribuciones de las semillas usados en la sección 4. Una vez dibujado el diagrama de Voronoi correspondiente, se realizaron 200 réplicas de la propagación de las grietas. Se debe hacer notar que el tiempo de dibujar un diagrama de Voronoi ahora es muy costoso. La Figura

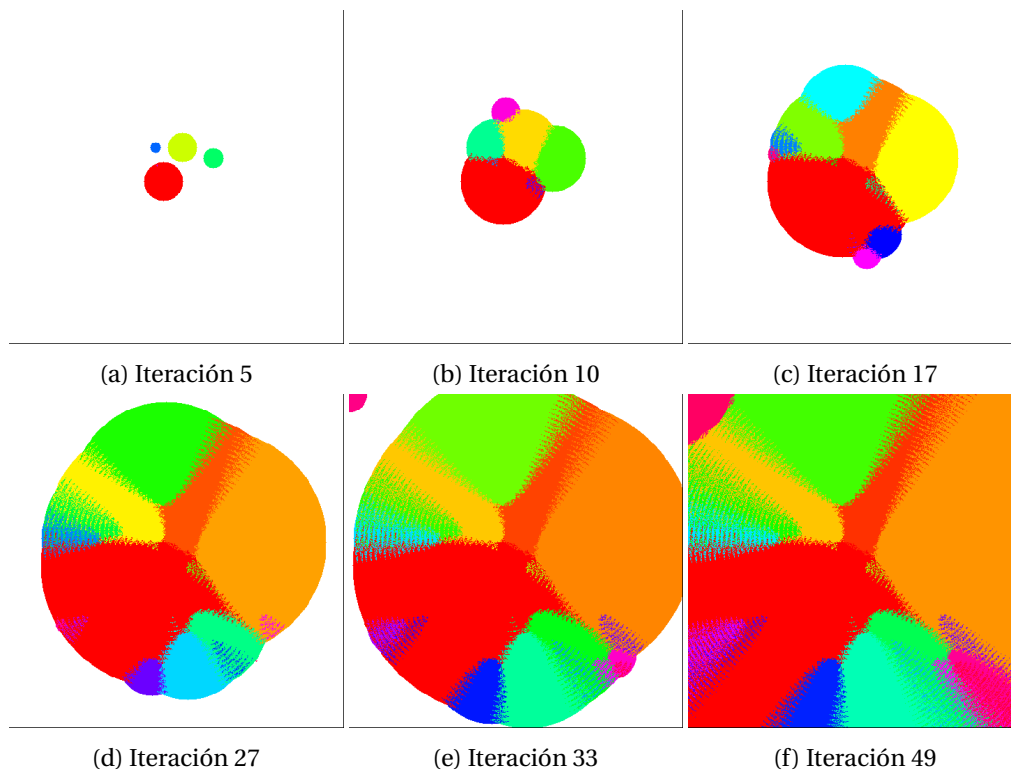


Figura 5.1: Ejemplo del crecimiento del diagrama de Voronoi cuando las semillas van apareciendo en el centro de la cuadrícula.

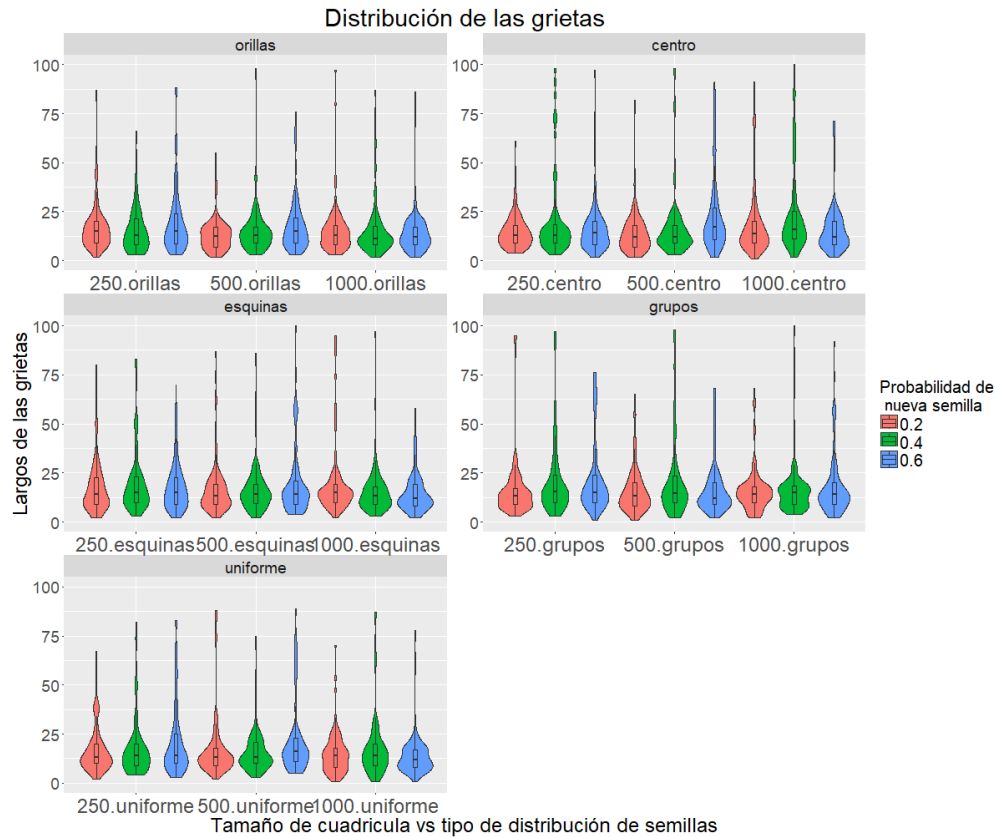


Figura 5.2: Distribución de la propagación de grietas en diagramas de Voronoi con crecimiento paulatino.

5.2 muestra la distribución de la formación de grietas en dependencia de los tres factores mencionados.

Como puede observarse a simple vista, no parece haber significancia de ningún factor. Los valores p de las pruebas de Kruskal y Wallis para determinar la significancia de cada factor aparecen en el Cuadro 5.1. Observe que hay muy poca significancia de los factores, lo cuál hace que digamos que los factores no tienen efecto sobre el tamaño de las grietas.

5.2. ¿QUÉ PASÓ?

Pues bien, como las pruebas estadísticas indican que haciendo crecer paulatinamente las celdas de Voronoi, no importa ni el tamaño, ni cómo ni cuándo se introduzcan las semillas, se hace la prueba para revisar la diferencia estadística entre las grietas generadas en un diagrama de Voronoi que se hizo de la forma usual, a aquellas generadas en un diagrama que creció paulatinamente. El diagrama de bigotes y violines correspondiente aparece en la Figura 5.3. La prueba estadística de Kruskal y Wallis indica una diferencia estadística con un valor p de 5.775×10^{-16} . Como puede observarse, utilizando un crecimiento paulatino las grietas son menores; sin embargo, las distribuciones de los largos de las grietas para ambos casos son

Factor	Valor p
Tamaño de la cuadrícula	0.02675
Distribución de semillas	0.001697
Probabilidad de nueva semilla	0.01657

Cuadro 5.1: Valores p para la significancia de los factores en el largo de las grietas.

muy similares.

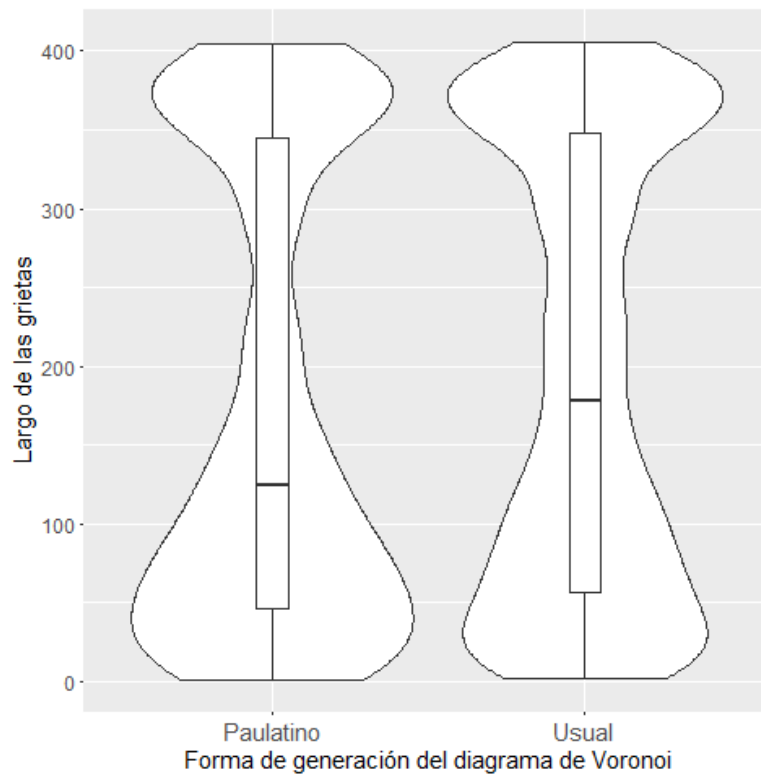


Figura 5.3: Comparación entre las distintas formas de generar el diagrama de Voronoi y su efecto en el crecimiento de las grietas.