

Preguntas

Jose Alejandro Montana Cortes

June 2020

1 Avances

Acá presento un resumen breve de lo que he estado explorando con respecto a la ultima reunión que tuvimos.

1.1 Acotando la probabilidad de tener L errores.

Como se habló la ultima reunión, mostrábamos que si considerábamos el operador

$$X_{12} \equiv \text{Tr}_{N/L} |\vec{n}_1\rangle\langle\vec{n}_2|,$$

este podía ser expandido en términos de operadores locales en L ($\gamma(\vec{x}, \vec{y})$). De forma que se tiene

$$\hat{X}_{12} = \sum_{\vec{x}, \vec{y}} f(\vec{x}, \vec{y}) \gamma(\vec{x}, \vec{y}). \quad (1)$$

Lográbamos demostrar que siempre que el vector de errores, definido como $e_{12} = \vec{n}_1 + \vec{n}_2 \pmod 2$, tuviera una cantidad de 1's mayor a L , esto implicaría inmediatamente que $f(\vec{x}, \vec{y}) = 0$. Con lo cual la primera pregunta que surge de esto era, ¿qué tan probable es que dos secuencias aleatorias de 0's y 1's de tamaño N tengan a lo más L errores?. Para contestar a esta pregunta primero redefinimos este problema en términos de unas variables aleatorias. Sea $X_i \theta_k$ una variable aleatoria que toma el valor 1 con probabilidad $p(\theta_k)$ y el valor 0 con probabilidad $1 - p(\theta_k)$, y sea $X = \sum_i X_i(\theta_k)$ la suma de estas variables aleatorias, o el numero de errores totales. Entonces nos preguntamos por la probabilidad de que el numero de errores sea menor o igual a L ($P(d \leq L)$). Para responder a esto hacemos uso de la desigualdad de Markov en un caso especial (Desigualdad de Chernoff).

$$P(X \geq L) \leq \frac{E(X)}{L},$$

en general esta desigualdad también cumple que (Chernoff)

$$P(X \geq L) = P(e^{\lambda X} \geq e^{\lambda L}) \leq \frac{E(e^{\lambda X})}{e^{\lambda L}}, \quad \forall \lambda > 0,$$

Sí adicional a esto hacemos uso de la independencia de las variables aleatorias $X_i(\theta_k)$, tenemos que el valor esperado puede escribirse como:

$$\begin{aligned} E(e^{\lambda X}) &= \prod_k E(e^{\lambda X(\theta_k)}) = \prod_k (1 + p(\theta_k)(e^\lambda - 1)) \geq \prod_k e^{p(\theta_k)(e^\lambda - 1)}, \\ &= e^{\sum_k p(\theta_k)(e^\lambda - 1)} = e^{\mu(e^\lambda - 1)} \end{aligned}$$

entonces,

$$P(X \geq L) \leq e^{\mu(e^\lambda - 1) - \lambda L}$$

Encontrando λ tal que minimice esta desigualdad, se tiene que $\lambda = \log(L/\mu + 1)$, con lo que se tiene entonces que

$$P(X \geq L) \leq e^{-N(\frac{L}{N}(\log(\frac{L}{\mu} + 1) - \frac{L}{N}))} \equiv e^{-NR(L/N)},$$

Donde $R(L/N)$ es una función que llamaremos acá "Rate Function". No obstante esta cota no ofrece una descripción acerca de la concentración de esta variable al rededor de un valor. Dado que esta variable es una variable de tipo de Bernoulli, es bien conocido que este tipo de variables aleatorias

tienden a concentrarse al rededor de su valor promedio, para ver esto consideremos las siguientes probabilidades.

$$P(X > (1 + \delta)\mu), \quad P(X < (1 - \delta)\mu).$$

Siguiendo los mismos pasos anteriores es fácil mostrar que

$$P(X > (1 + \delta)\mu) < \left(\frac{e^\delta}{(1 + \delta)(1 + \delta)} \right)^\mu, \quad (2)$$

y

$$P(X < (1 - \delta)\mu) < \left(\frac{e^{-\delta}}{(1 - \delta)(1 - \delta)} \right)^\mu \quad (3)$$

Lo cual muestra que este tipo de variables tiende a estar muy concentrado con respecto a su valor promedio. Dicho esto, resulta de gran utilidad calcular este valor esperado, para esto recordamos que cada valor en nuestras secuencias toma un valor de 1 con probabilidad $n(\theta_k)$ y 0 con $1 - n(\theta_k)$. Explícitamente, esta probabilidad viene de la distribución Fermi-dirac,

$$n(\theta_k) = \frac{1}{1 + e^{\beta(\Omega(\theta_k) - \Omega^*)}}, \quad (4)$$

con $\Omega(\theta_k)$ el espectro, que para nuestro caso es

$$\Omega(\theta_k) = \sqrt{(\lambda - \cos \theta_k)^2 + \gamma^2 \sin^2 \theta_k},$$

con λ y γ los parámetros para el modelo XY y $\sqrt{\gamma^2(\gamma^2 - 1 + \lambda^2)/(\gamma^2 - 1)}$, el mínimo de energía. De esta forma podemos calcular una expresión para la probabilidad de que 2 secuencias tengan un error en la posición k .

$$p(\theta_k) = n(\theta_k)^2 + (1 - n(\theta_k))^2 = \frac{1}{1 + \cosh(\beta(\Omega(\theta_k) - \Omega^*))}, \quad (5)$$

y por ende la probabilidad de que dos secuencias sean iguales en la posición k es

$$1 - p(\theta_k) = \frac{1}{1 + \operatorname{sech}(\beta(\Omega(\theta_k) - \Omega^*))}. \quad (6)$$

Por ende el valor esperado de los errores, $X = \sum_k X(\theta_k)$, está dado por:

$$\mu = \sum p(\theta_k) \underset{N \rightarrow \infty}{=} \frac{N}{2\pi} \oint d\theta \frac{1}{1 + \cosh(\beta(\Omega(\theta) - \Omega^*))}. \quad (7)$$

En las figuras 1 2 se muestra el comportamiento de la distribución de probabilidad asociada a

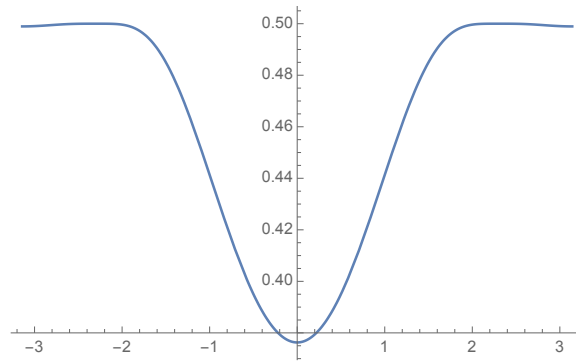


Figure 1: Distribución de probabilidad de errores en función de la posición de la cadena para $\beta = 1$.

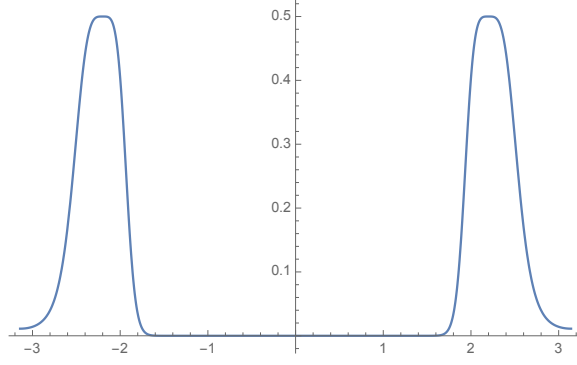


Figure 2: Distribución de probabilidad de errores en función de la posición de la cadena para $\beta = 80$.

Con el fin de ver el comportamiento del valor esperado en función de la temperatura, expandimos (6) en función de $\beta \rightarrow 0$.

$$\begin{aligned}\mu &\approx \frac{N}{2\pi} \left(\frac{2\pi}{2} - \beta^2 \int_{-\pi}^{\pi} \frac{(\Omega(\theta) - \Omega^*)^2}{8} d\theta \right) \\ &= \frac{N}{2} - \frac{N\beta^2}{16} \int_{-\pi}^{\pi} (\Omega(\theta) - \Omega^*)^2 d\theta,\end{aligned}\quad (8)$$

es decir que cuando estemos tomemos una temperatura alta ($T \rightarrow \infty$, el valor esperado μ tenderá a $N/2$. Para el caso en que $\beta \rightarrow \infty$, podemos aproximar la integral (7) como

$$\mu \rightarrow \frac{N}{2\pi} \oint e^{-\beta(\Omega(\theta) - \Omega^*)} d\theta \rightarrow 0 \quad (9)$$

Con el fin de mejorar las cota encontradas en (2) y (3). Podemos hacer uso de la desigualdad de Hoeffding. Llamando

$$I = \frac{1}{2\pi} \oint \frac{d\theta}{1 + \cosh(\beta(\Omega(\theta) - \Omega^*))}$$

Es posible mostrar que

$$P(N(I - \varepsilon) \leq X \leq N(I + \varepsilon)) \geq 1 - 2e^{-2\varepsilon^2 N} \quad (10)$$

Adicional a esto tenemos que la varianza está dada por

$$\sigma^2 = \sum_k p(\theta_k)(1 - p(\theta_k)) \quad \xrightarrow{N \rightarrow \infty} \quad \frac{N}{2\pi} \oint \frac{\cosh(\beta(\Omega(\theta) - \Omega^*))}{(1 + \cosh(\beta(\Omega(\theta) - \Omega^*)))^2} d\theta. \quad (11)$$

De esta forma tenemos entonces que al tomar el limite en que $\beta \rightarrow 0$

$$\sigma^2 \approx \frac{N}{4} - \frac{N}{128\pi} \beta^4 \int_{-\pi}^{\pi} (\Omega(\theta) - \Omega^*)^4 d\theta \quad (12)$$

y cuando $\beta \rightarrow \infty$, tenemos que el comportamiento es similar al de (9)

1.2 Acotando las transformaciones O y V

Como discutimos la anterior vez, queremos probar que incluso en el caso en el que tuviéramos dos secuencias para las cuales el numero de errores fuera menor a L , el producto cruzado debe de ser de igual forma pequeño. Entonces para esto podemos estudiar

$$\|X_{12}\|_2^2 = \text{Tr } X_{12} X_{12}^\dagger \quad (13)$$

Como lo se discutió en la ultima sesión, esta norma puede ser escrita como:

$$\|X_{12}\|_2^2 = \frac{1}{2^L} \sum_{\vec{x}, \vec{y}} (O_{\vec{x}\vec{x}'} O_{\vec{x}\vec{x}''}) (V_{\vec{y}\vec{y}'} V_{\vec{y}\vec{y}''}), \quad (14)$$

no obstante, veíamos que al interpretar estas secuencias \vec{x} como p -formas, se tenía que la si se quería que esto transformara como una p -forma y que además los operadores transformaran como variables

de Grassmann, es necesario que estas transformaciones se efectuaran por medio de un determinante, con esto teníamos entonces que

$$\underbrace{\sum_{\vec{x}}}_{\text{Suma sobre p-formas}} (O_{\vec{x}\vec{x}'} O_{\vec{x}\vec{x}''}) = \det \left[(O^T \Pi_L O) |_{\vec{x}', \vec{x}''} \right],$$

en donde Π_L es el proyector al espacio L de la cadena y $|_{\vec{a}, \vec{b}}$ representan el menor sobre \vec{a}, \vec{b} . Como primera aproximación se efectuó un análisis de que tan grande es este determinante en términos de los grados de las p -formas, o como se conoce en teoría de información, el peso de la secuencia. Con esto llegábamos entonces a plantear una posible cota en términos de la cota de Haddamard para el determinante. Así se tenía entonces que

$$\|X_{12}\|_2^2 \leq \frac{1}{2^L} \sum_{w_1, w_2} \binom{d}{w_1}^2 \binom{n-d}{w_2}^2 \left[\frac{2}{N} (w_1 + w_2) \right]^{w_1 + w_2} \left[\frac{2}{N} (d - w_1 + w_2) \right]^{d - w_1 + w_2}, \quad (15)$$

con,

$$\begin{aligned} 0 &\leq w_1 \leq d \\ 0 &\leq w_2 \leq N - d \\ 0 &\leq w_1 + w_2 \leq L \\ 0 &\leq d - w_1 + w_2 \leq L, \end{aligned} \quad (16)$$

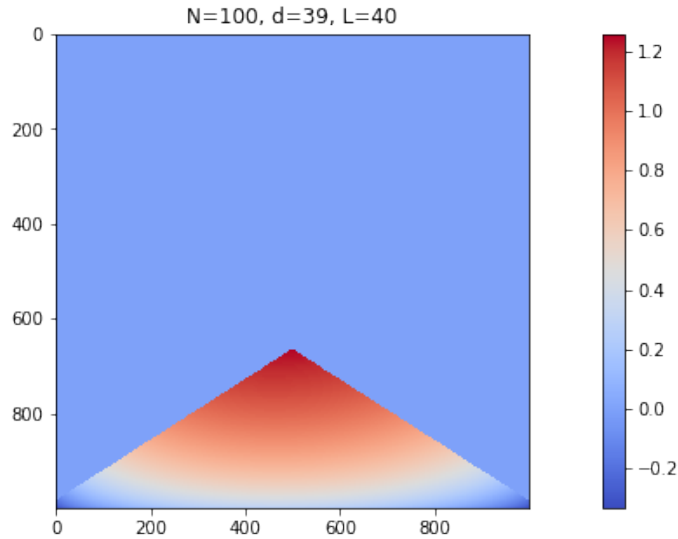
y d el numero de errores. Si suponemos que esto sigue una ley de grandes desviaciones, tenemos entonces que

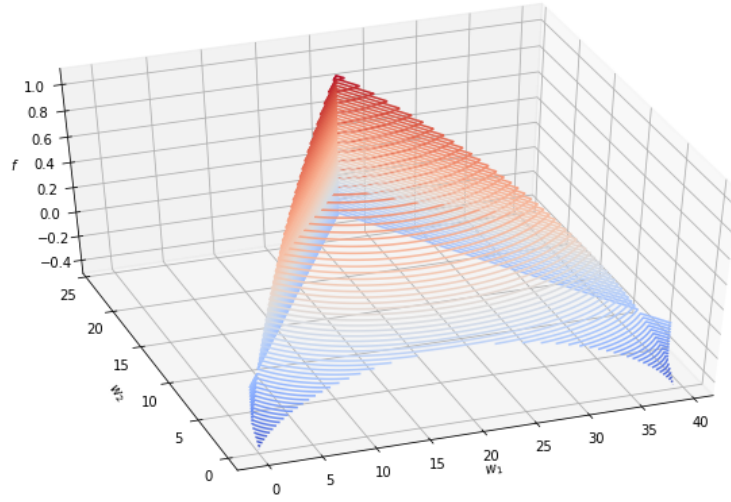
$$\|X_{12}\|_2^2 \leq e^{Nf(w_1, w_2, L, d, N)}, \quad (17)$$

y

$$\begin{aligned} f(w_1, w_2, L, d, N) = & 2 \frac{d}{N} H_2 \left(\frac{w_1}{2} \right) + 2 \left(1 - \frac{d}{N} \right) H_2 \left(\frac{w_2}{N-d} \right) + \frac{w_1 + w_2}{N} \log_2 \left(2 \frac{w_1 + w_2}{N} \right) \\ & + \frac{d - w_1 + w_2}{N} \log_2 \left(2 \frac{d - w_1 + w_2}{N} \right) - \frac{L}{N} \end{aligned} \quad (18)$$

Desafortunadamente como se muestra en las figuras ?? y ?? el comportamiento de f sobre la región descrita en (16) hace que no sea posible asegurar que existirá una cota pequeña para distintos valores de N, L, d





Es por esto que esta forma de ver el problema no resulta muy útil. No obstante, es posible hacer un experimento numérico y ver que el determinante de las matrices $O^T \Pi_L O$ resulta ser mucho más pequeño de lo esperado.

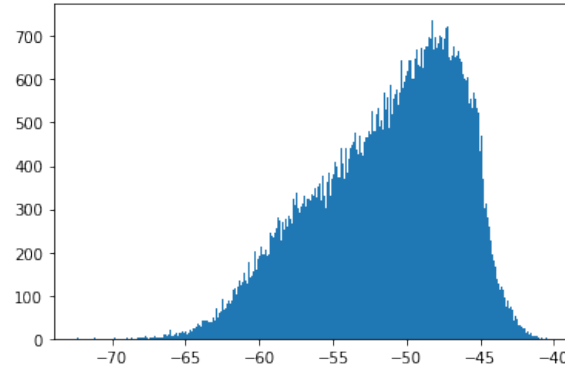


Figure 3: Histograma del determinante de los menores de la matriz $O^T \Pi_L O$, para un tamaño de la cadena de $N = 1000$ y $L = 20$

Como se muestra en la figura 3 al tomar el logaritmo del valor absoluto del determinante de distintos menores, se puede evidenciar que este determinante es bastante pequeño, si variamos L en para un N fijo podemos construir una superficie con estos Histogramas y ver que para varios valores de L se tiene este mismo comportamiento. En la figura 4 se ilustra este hecho. Con esto en mente, se tendría entonces que $\|X_{12}\|_2^2$ también sería bastante pequeño. Pero para entender por que ocurre esto, es necesario encontrar una forma explicita para $O^T \Pi_L O$, esto puede hacerse haciendo uso de la relación con el Kernel de Dirichlet, pues se tiene que las filas de O pueden ser escritas como modos de funciones cosenos de la forma:

$$u^k(x) = \sqrt{\frac{2}{N}} \cos \theta_k x, \quad (19)$$

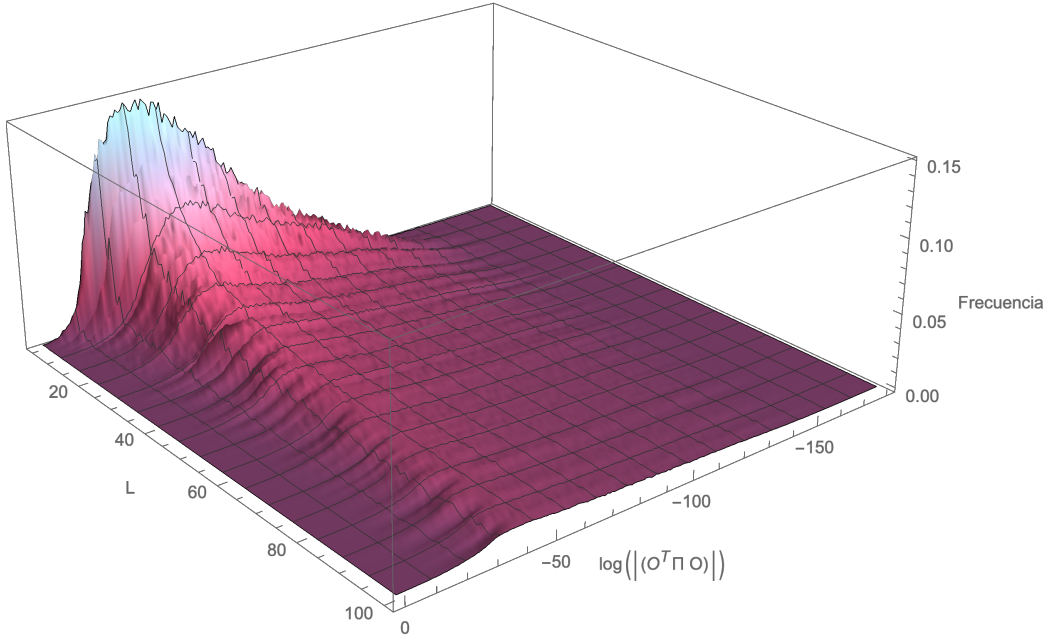


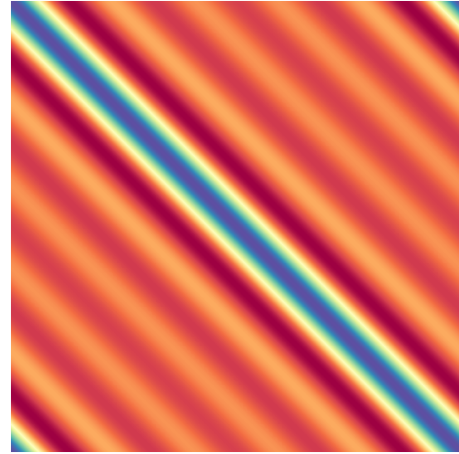
Figure 4: Superficie formada por los histogramas variando L y dejando N fijo

dado que en nuestro interés es hallar la forma de un producto de dos de estas funciones, tenemos entonces que

$$\sum_{\vec{x}} O_{\vec{x}, \vec{x}'} O_{\vec{x}, \vec{x}''} \equiv A \rightarrow A_{i,j} = \sum_{x=-\frac{L-1}{2}}^{\frac{L-1}{2}} u^i(x) u^j(x), \quad (20)$$



(a) Matriz $O^T \Pi_L O$, con $N = 8000$ y $L = 30$



(b) Matriz hecha por medio del Kernel de Dirichlet $1/N D_L(\frac{2\pi}{N}(\vec{x}' - \vec{x}''))$, con $N = 8000$ y $L = 30$

Figure 5: Como se aprecia en las figuras las dos matrices son idénticas, para este caso se tiene un rms de 10^{-16}

de donde se deduce que¹:

$$= \frac{1}{N} D_L(\theta_j - \theta_k), \quad (21)$$

¹Acá se hace uso del hecho de que

$$\cos \theta_i x \cos \theta_j x = \frac{1}{2} (\cos(x(\theta_j - \theta_k)) + \cos(x(\theta_j + \theta_k)))$$

donde $D_L(x)$ corresponde al kernel de Dirichlet dado por

$$D_L(x) = \frac{\sin\left(x\frac{L}{2}\right)}{\sin\left(\frac{x}{2}\right)}.$$

Con esto tenemos entonces que al escribir podemos escribir nuestra matriz $O^T \Pi_L O$ en términos de \vec{x}', \vec{x}'' como:

$$O^T \Pi_L O|_{\vec{x}', \vec{x}''} = \frac{1}{N} D_L\left(\frac{2\pi}{N}(\vec{x}' - \vec{x}'')\right). \quad (22)$$

Con lo cual tenemos es que encontrar una cota para el determinante de este menor. En la figura 5 se muestran las gráficas del valor de $O^T \Pi_L O$ y de $\frac{1}{N} D_L\left(\frac{2\pi}{N}(\vec{x}' - \vec{x}'')\right)$. De forma particular, es posible hacer uso de la Cota de Hadamard para el caso de estos menores. Recordamos que estos menores serán de tamaño máximo L , a esto le añadimos el hecho de que el valor máximo que puede tomar el Kernel de Dirichlet es L , de con lo cual tenemos entonces que

$$\det(O^T \Pi_L O) \leq \left(\frac{L}{N}\right)^L L^{L/2} = \left(\frac{L^{3/2}}{N}\right)^L \ll 1, \quad \text{para } L < N, \quad (23)$$

No obstante, dado que tenemos que la matriz $(O^T \Pi_L O)$ es Toeplitz, estamos tratando con el determinante del menor de una matriz que tiene esta particularidad. Por ende es posible hacer uso de un caso particular de la dualidad Frobenius-Schur para obtener una cota más ajustada sobre este tipo de matrices. En particular es posible utilizar una generalización del teorema de Szegő, el cual lleva el nombre de Bump, Diaconis. Dentro de este teorema se encuentra una forma asintótica para el determinante de menores de matrices Toeplitz. De forma sorprendente parece que el carácter del grupo simétrico $(U(n))$ aparece en esta formula.

Si λ y μ son particiones de longitud n , se puede mostrar que

$$D_{n-1}^{\lambda, \mu}(f) = \det(d_{\lambda_i - \mu_j - i + j}), \quad (24)$$

Siendo d_i los elementos de la matriz Toeplitz, es posible mostrar entonces que la expresión (24) coincide con el menor de una matriz Toeplitz “Grande”.

Teorema (Heine, Szegő, Bump, Diaconis):

Sea $f \in L^1(\mathbb{T})$ dado por $f(t) = \sum_{n=-\infty}^{\infty} d_n t^n$. y sean λ y μ particiones de longitud n . Se define una función $\Phi_{n,f}(g)$ sobre $U(n)$, como $\Phi_{n,f}(g) = \prod_{i=1}^n f(t_i)$, donde t_i son los eigen-valores de $g \in U(n)$. Entonces

$$D_{n-1}^{\lambda, \mu}(f) = \int_{U(n)} \Phi_{n,f}(g) \overline{\chi_\lambda(g)} \chi_\mu(g) dg, \quad (25)$$

Con χ_i el carácter de $U(n)$. Particularmente si λ y μ son las particiones triviales, la identidad Heine-Szegőes recuperada.

Una particularidad que encontré es que el para el caso del grupo $SU(2)$, el carácter del grupo está dado por el kernel de de Dirichlet, quizás esto puede ser una coincidencia pero creo que de igual forma esto nos podría ayudar a calcular una mejor cota.

2 Notas:

- Dentro de esta ultima parte que he estado estudiando aunque muy interesante me ha parecido que tiene una serie de conceptos que nunca he tocado antes, con lo cual se me ha complicado un poco entender que está detrás de todo esto como para tener una idea de como poder utilizar este resultado para nuestros cálculos.
- Dentro de la formulación que teníamos para ver el peso de las secuencias me surgieron dudas acerca de como es que podemos ver esto exactamente como un P-forma, aunque sigo el hilo de ideas me parece un poco “mágica” esta forma de ver el problema.

- Para el caso de la función de probabilidad me han surgido bastantes dudas con algo que aunque puede ser una bobada, parece que si le ponemos la energía menos su valor de Gap, tenemos que entonces existen regiones de la cadena para las cuales la probabilidad de haber un error es siempre de $1/2$. Esto me parece un poco extraño ya que es como si entonces existieran regiones sobre la cadena que tienen la propiedad de que es más probable que esté desordenado.
- Lo otro que tenía de duda es acerca del Paper que me mandó, pues aunque no he todas las cuentas exhaustivamente, al tener que ver con códigos de distancia mínima relacioné que esto tendría que ver con lo primero que hablamos en la ultima sesión acerca de ver que tan grande puede ser ese espacio para el cual se cumple que estos códigos se puedan corregir y que se de origen a lo que llamábamos el espacio de Hilbert “Congelado”. No obstante aún tengo unas dudas con respecto a esto pues en principio lo que queríamos ver era esta unión de espacios (que por cierto no es necesariamente espacio vectorial), cómo se relacionan estos exponentes de error de códigos binarios con esto que queremos tratar de estudiar.