Talagrand's Convex Hull Concentration Inequality with Applications

Jon-Paul Mastrogiacomo

Fall 2018

Contents

1 Introduction

Consider $\Omega = \prod_{i=1}^{n} \Omega_i$ where each Ω_i is a probability space and Ω has the product measure. Let $A \subseteq \Omega$.

We would like to say something about the probability of a random variable in Ω landing in A. But we have not specified a measure on any of our Ω_i spaces so we have little information to work with. Our main tool will be the so-called convex distance $\rho(A, x)$ that measures the distance from a point x to A in a clever way. This will be used in the following theorem called Talagrand's inequality

$$Pr[A]Pr[A_t^c] \leq e^{-t^2/4}$$

where the convex distance is used to define our A_t set. This is a manifestation of the concentration of measure phenomenon and in some cases it is a direct improvement on other concentration inequalities. Talagrand's inequality can tell us about probabilistic quantities of many variables. For example under certain assumptions it can show that if a function does not depend too much on any one variable, then it is concentrated around its expectation. We begin by trying to understand how our distance function behaves.

2 The Convex Distance

2.1 First Definition

Consider $x = (x_1, \ldots, x_n) \in \Omega$. First we introduce the Hamming Distance for some other $y = (y_1, \ldots, y_n) \in \Omega$:

$$d(x, y) = \sum_{i=1}^{n} 1_{\{x_i \neq y_i\}}$$

This is just the number of elements that differ between strings or vectors. Now for any $\alpha \in \mathbb{R}^n$ we define the α-weighted Hamming Distance:

$$d_\alpha(x, y) = \sum_{i=1}^{n} \alpha_i 1_{\{x_i \neq y_i\}}$$

Now for our $A \subseteq \Omega$ we take the shortest "distance" to the set:

$$d_\alpha(A, x) = \inf_{y \in A} d_\alpha(x, y)$$

We define the convex distance by maximizing the sum over all possible choices of α with $|\alpha| = \sqrt{\sum \alpha_i^2}$:

$$\rho(A, x) = \sup_{|\alpha|=1} d_\alpha(A, x)$$

Alternatively let the convex distance ρ be the least real number such that for some α with $|\alpha| = 1$ there exists $y \in A$ with

$$\sum_{i:x_i \neq y_i}^{n} \alpha_i \leq \rho(A, x)$$

These are equivalent statments. Lastly for any real $t \geq 0$ we define the set,

$$A_t = \{x \in \Omega : \rho(A, x) \leq t\}$$

and A_t^c to be its compliment. Note that when $x \in A$ we can select $y = x$ so that $A_0 = A$. The α-weighted hamming distance is often used in computer science to analyse high dimensional data, which is what we are doing here: reasoning about how close x is to A.

2.2 Second Definition and Equivalence

Define $U(A, b)$ to be the set of $s = (s_1, \ldots, s_n) \in \{0, 1\}^n$ with the property that there exists $y \in A$ such that

$$x_i \neq y_i \implies s_i = 1$$

This is some set in the binary hypercube that we can think of as representing the possible paths from x to A. For technical reasons $s_i = 1$ does not mean $x_i \neq y_i$. If some $u \in U(A, x)$ has $u_i = 0$ it means that $x_i = y_i$ and we do not have to adjust these coordinates. We concern ourselves with the u_i such that $u_i = 1$. Note that now $\rho(A, x)$ is the least real for all $u \in U(A, x)$ and $\alpha \in \mathbb{R}^n$ with $|\alpha| = 1$ such that $\alpha \cdot u \leq \rho(A, x)$.

Next is a definition of the convex hull that we will rely on for the duration of this report. The convex hull of a finite set B with N points is the intersection of all convex sets containing B, or equivalently, the unique minimal convex set containing B. The convex hull of $b_1, b_2, \ldots, b_N \in B$ is given by

$$\left\{ \sum_{i=1}^{N} \lambda_j b_j : \lambda_j \geq 0 \text{ and } \sum_{i=1}^{N} \lambda_j = 1 \right\}$$

Now we define $V(A, x)$ as the convex hull of $U(A, x)$. That is, the unique minimal convex set containing $U(A, x)$. The following result provides a new way to think about this distance function ρ that sits at the heart of Talagrand's Inequality.

Theorem.

$$\rho(A, x) = \min_{v \in V(A,x)} |v|$$

Proof. Suppose $v \in V(A, x)$ achieves this minimum. We can draw a hyperplane through v that entirely separates $V(A, x)$ from the origin. For all $s \in V(A, x)$ which is convex, we know $s \cdot v \geq v \cdot v$. If we set $\alpha = \frac{v}{|v|}$ then for all $s \in U(A, x) \subseteq V(A, x)$ we have $\rho(A, x) \geq s \cdot \alpha \geq v \cdot \frac{v}{|v|} = |v|$

Conversely, take any α with $|\alpha| = 1$ and again assume v assumes this minimum. Clearly $\alpha \cdot v \leq |v|$. From the definition of convex hull and since $v \in V(A, x)$ we can write $v = \sum \lambda_i t_i$ for some $t_i \in U(a, x)$ where all $\lambda_i \geq 0$ and $\sum \lambda_i = 1$. Then

$$|v| \geq \alpha \cdot v = \alpha \cdot \sum \lambda_i t_i = \sum \lambda_i (\alpha \cdot t_i)$$

hence, $\alpha \cdot t_i \leq |v|$ for any t_i we choose. But we know $\rho(A, x)$ is the least real such that $\rho(A, x) \geq \alpha \cdot t_i$ so finally $\rho(A, x) \leq |v|$. □

One way to picture the convex distance is to imagine the point x moving to the origin and the set A collapsing into to binary hypercube. Then we just measure the Euclidean distance from the origin to the convex hull of our new A set. In general, the first definition is better for applications as we often have our α_i values explicitly. The second version is more of a theoretical tool and we use it exclusively to prove Talagrand's inequality.

3 Talagrand's Inequality

The first theorem holds the insight and nuance of Talagrand's inequality. We will prove it by a beautiful induction argument that makes use of the convex distance we have been working with. However it does not look like a traditional concentration inequality. After an easy application of Chebyshev's inequality we arrive at the more pleasing form in theorem two. Throughout the rest of the report we will denote $Pr[X \in A]$ by simply $Pr[A]$ where X is a random variable in Ω. We will also suppress the measure notation from our integrals. Since measures on the sets Ω_i are unrelated, we can just apply Fubini's theorem by integrating on each separately one at a time.

Theorem 1.

$$\int_\Omega \exp\left[\frac{1}{4}\rho^2(A, x)\right] \leq \frac{1}{Pr[A]}$$

Theorem 2 (Talagrand's Inequality)**.**

$$Pr[A]Pr[A_t^c] \leq e^{-t^2/4}$$

Proof. First we fix A and consider $X = \rho(A, x)$ as a random variable. After a simple application of Chebyshev's Inequality we have,

$$Pr[A_t^c] = Pr[X \geq t] = Pr[e^{X^2/4} \geq e^{t^2/4}] \leq E[e^{X^2/4}]e^{-t^2/4}$$

Now we use the previous theorem so that

$$E[e^{X^2/4}] = \int_\Omega e^{X^2/4} \leq \frac{1}{Pr[A]} \implies Pr[A_t^c] \leq \frac{e^{-t^2/4}}{Pr[A]}$$

□

3.1 Proof of Theorem 1

We proceed by induction on the dimension n. For $n = 1$ we only have to vary one coordinate, so $\rho(A,x) = 1$ if $x \notin A$ and zero otherwise so that

$$\int_\Omega \exp\left[\frac{1}{4}\rho^2(A,x)\right] = \int_A \exp\left[\frac{1}{4}\rho^2(A,x)\right] + \int_{A^c} \exp\left[\frac{1}{4}\rho^2(A,x)\right] = Pr[A] + Pr[A^c]e^{\frac{1}{4}} \leq \frac{1}{Pr[A]}$$

since $c + (1-c)e^{\frac{1}{4}} \leq c^-1$ for $0 < c \leq 1$.

Assume the result holds for n. Let OLD $= \prod_{i=1}^n \Omega_i$ and NEW $= \Omega_{n+1}$ so that $\Omega =$OLDxNEW. Any $z \in \Omega$ can now be uniquely written as $z = (x,\omega)$ with $x \in$ OLD and $\omega \in$ NEW. Set

$$B = \{x \in \text{OLD} : (x,\omega) \in A \text{ for some } \omega \in A\}$$

and for any $\omega \in$ NEW set

$$A_\omega = \{x \in \text{OLD} : (x,\omega) \in A\}$$

Given $z = (x,\omega) \in \Omega$ we can move to $A \subseteq \Omega$ in two different ways. If we are willing to change ω, we reduce our problem to moving from x to B. So we change x until we find a suitable ω so that (x,ω) lies in A. If we do not want ω to change, we need to move from x to A_ω. Thus

$$s \in U(B,x) \implies (s,1) \in U(A,(x,\omega))$$

and

$$t \in U(A_\omega,x) \implies (t,0) \in U(A,(x,\omega))$$

Note that this is where the condition $x_i = y_i \nRightarrow s_i = 0$ or equivalently $s_i = 1 \nRightarrow x_i \neq y_i$ is important. If we are moving from $t \in U(A_\omega,x)$ to $U(A,(x,\omega))$ we can guarantee that the last coordinate lies in A since we already have the ω'th coordinate in the set, hence $\omega = 0$. But if we are moving from $s \in U(B,x)$, there are some values of the ω'th coordinate that lie in A and some that do not. So we need $\omega = 1$ to be flexible either way.

We take the convex hulls, so if $s \in V(B,x)$ and $t \in V(A_\omega,x)$ then $(s,1)$ and $(t,0)$ are in $V(A,(x,\omega))$. Here we are viewing $U(B,x)$, $U(A_\omega,x)$, and $U(A,(x,\omega))$ as subsets of their convex hulls. By the above definition of a convex hull

$$((1-\lambda)s + \lambda t, 1-\lambda) = ((1-\lambda)s + \lambda t, (1-\lambda)\cdot 1 + \lambda \cdot 0) \in V(A,(x,\omega))$$

Since ρ is the shortest of all distances to the convex hull

$$\rho(A,(x,\omega)) \leq |(1-\lambda)s + \lambda t, 1-\lambda|$$

By convexity of the Euclidean norm and the square function (ie $u \to |u^2|$ is convex), we can apply the definition of a convex function and the triangle inequality so that

$$\begin{aligned}\rho^2(A,(x,\omega)) &\leq |(1-\lambda)s + \lambda t|^2 + (1-\lambda)^2 \\ &\leq (1-\lambda)|s|^2 + \lambda|t|^2 + (1-\lambda)^2\end{aligned}$$

Remember $s \in V(B,x)$ and $t \in V(a_\omega, x)$ so if we pick s and t to be the points with minimal norm $|s| = \rho(B,x)$ and $|t| = \rho(A_\omega, x)$. We are slightly abusing notation since s and t were originally defined in the sets of discreet points but now we treat them as members of the convex hulls. This yields the critical inequality

$$\rho^2(A,(x,\omega)) \leq (1-\lambda)\rho^2(B,x) + \lambda\rho^2(A_\omega,x) + (1-\lambda)^2$$

Remember $x \in \text{OLD}$ so we can fix ω and bound the integral over all possible values of x

$$\begin{aligned}\int_{\text{OLD}} \exp\left[\frac{1}{4}\rho^2(A,(x,\omega))\right] &\leq \int_{\text{OLD}} \exp\left[\frac{1}{4}\Big((1-\lambda)\rho^2(B,x) + \lambda\rho^2(A_\omega,x) + (1-\lambda)^2\Big)\right] \\ &\leq e^{(1-\lambda)^2/4}\int_{\text{OLD}}\left(\exp\left[\frac{1}{4}\rho^2(B,x)\right]\right)^{1-\lambda}\left(\exp\left[\frac{1}{4}\rho^2(A_\omega,x)\right]\right)^{\lambda}\end{aligned}$$

By Hölder's Inequality (see section 5.1) this is at most

$$e^{(1-\lambda)^2/4}\left(\int_{\text{OLD}}\exp\left[\frac{1}{4}\rho^2(B,x)\right]\right)^{1-\lambda}\left(\int_{OLD}\exp\left[\frac{1}{4}\rho^2(A_\omega,x)\right]\right)^{\lambda}$$

We apply the induction hypothesis to each integral separately. Therefore the previous statement is at most

$$\begin{aligned}e^{(1-\lambda)^2/4}\left(\frac{1}{Pr[B]}\right)^{1-\lambda}\left(\frac{1}{Pr[A_\omega]}\right)^{\lambda} &= e^{(1-\lambda)^2/4}\left(\frac{1}{Pr[B]}\right)\left(\frac{1}{Pr[B]}\right)^{-\lambda}\Big(Pr[A_\omega]\Big)^{-\lambda} \\ &= \frac{1}{Pr[B]}e^{(1-\lambda)^2/4}\gamma^{-\lambda}\end{aligned}$$

where $\gamma = Pr[A_\omega]/Pr[B] \leq 1$ because $A_w \subseteq B$. Now we apply the argument in section 5.2 and find

$$\int_{\text{OLD}}\exp\left[\frac{1}{4}\rho^2(A,(x,\omega))\right] \leq \frac{1}{Pr[B]}(2-\gamma) = \frac{1}{Pr[B]}\left(2 - \frac{Pr[A_\omega]}{Pr[B]}\right)$$

Now if we integrate over ω, $Pr[B]$ does not change because the set B does not change as we vary ω. It is fixed at its definition by considering all $\omega \in \text{NEW}$. However as ω changes, A_ω certainly changes. As we integrate we are adding ω slices of A for which all

$x \in$ OLD lie in A. So we get back all of A by integrating! This is just Fubini's theorem. Therefore considering $\omega \in$ NEW

$$\int_{\text{NEW}} \int_{\text{OLD}} \exp\left[\frac{1}{4}\rho^2(A,(x,\omega))\right] \le \frac{1}{Pr[B]}\left(2 - \frac{Pr[A]}{Pr[B]}\right) = \frac{1}{Pr[A]} x(2-x) \le \frac{1}{Pr[A]}$$

where $x = Pr[A]/Pr[B] \in [0,1]$. But we know $x(2-x) \le 1$, completing the induction and hence the theorem.

4 Bin Packing

This is an important problem in applied mathematics and computer science and Talagrand's inequality offers an insightful way to approach the problem. The function that describes the minimum number of bins is so complicated that it cannot be studied explicitly. Instead we have to use concentration inequalities to bound the probability that a random variable deviates from is expectation (or median). In this section we will compare Azuma's inequality to the more sophisticated Talagrand's inequality, however we could also use it to get an explicit concentration.

Consider numbers $x_1, x_2, \dots, x_n \in [0,1]$ packed into bins of size 1 based on their sums. So 0.4 and 0.5 can be packed in the same bin, but 0.6 and 0.7 cannot. Let f(x) denote the minimum number of bins required. Consider the random variable f(X) where $X_1, X_2, \dots, X_n$ are independent random variables taking values in [0,1].

We begin with this well known result but forgo a proof in this report.

Theorem (Azuma's Inequality)**.** *Let $X_1, X_2, \dots, X_n$ be independent random variables. Let $f(X_1, X_2, \dots, X_n)$ obey the following stability condition*

$$|f(x_1, \dots, x_i, \dots, x_n)| - |f(x_1, \dots, x_i', \dots, x_n)| \le a_i$$

for all $i \le n$ and for constants $a_1, a_2, \dots, a_n$. Then we have

$$Pr(f - Ef \ge t) \le \exp\left(-\frac{t^2}{2\sum_{i=1}^n a_i^2}\right)$$

In the case of bin packing, changing any one x_i can only add at most one bin. So if it cannot fit in any existing bin, it gets one of its own. So the stability condition of Azuma's inequality implies $a_i = 1$ for all $i \le n$. And we get the following bound on the probability that the number of bins deviates from its average

$$Pr(f(X) - Ef(X) \ge t) \le \exp\left(-\frac{t^2}{2n}\right)$$

If we want this probability to be less than some δ

$$\delta = \exp\left(-\frac{t^2}{2n}\right) \implies t = \sqrt{2n \log\frac{1}{\delta}}$$

We can restate the amount f deviates from its mean as follows

$$Pr\left(f(X) - Ef(X) \leq \sqrt{2\log\frac{1}{\delta}} \cdot \sqrt{n}\right) \geq 1 - \delta$$

Now we pivot to Talagrand's Inequality.

We need one property of $f(x)$ which represents the minimum number of bins required. For any $x \in [0,1]^n$

$$f(x) \leq 2\sum_{i=1}^{n} x_i + 1$$

since at worst, each x_i is slightly more than a half and each gets its own bin. This brings $\sum_{i=1}^{n} x_i$ to slightly more than n. The +1 accounts for the case that all of our values fit in one bin. Using this fact, for any $x, y \in [0,1]^n$ we get

$$f(x) \leq f(y) + \sum_{i:x_i \neq y_i} x_i + 1$$

since the x_i such that $x_i \neq y_i$ is like its own new packing and we can use the previous result.

Now let $\alpha = \alpha(x) \in [0,\infty)^n$ be the unit vector $x/||x||$. We have

$$\sum_{i:x_i \neq y_i} x_i = ||x|| \sum_{i:x_i \neq y_i} \alpha_i = ||x|| d_\alpha(x,y)$$

where d_α is the α-weighted Hamming Distance.

Now let M be the median of f and set $A_M = \{y : f(y) \leq M\}$. By definition of the convex distance ρ and the above argument, for each $x \in [0,1]^n$ there exists $y \in A_M$ such that

$$f(x) \leq f(y) + 2\sum_{i:x_i \neq y_i} x_i + 1 \leq M + 2||x||\rho(A_M, x) + 1 \tag{1}$$

From the statement of Talagrand's Inequality we can conclude

$$\begin{aligned} Pr\Big(\rho(A_M, x) \geq t\Big) &\leq \frac{1}{Pr(A_M)} e^{-t^2/4} \\ &\leq 2e^{-t^2/4} \end{aligned}$$

since the probability of lying on one side of the median is 1/2. If we set this probability to be within the same bound δ as above we have

$$\delta = 2e^{-t^2/4} \implies t = \sqrt{4\log\frac{2}{\delta}}$$

We can adjust what we have as follows

$$
\begin{aligned}
&Pr\Big(\rho(A_M, x) \leq t\Big) \geq 1 - \delta \\
\implies Pr\Big(Mf(X) + 2||X||\rho(A_M, x) + 1 \leq Mf(X) + 2||X||t + 1\Big) &\geq 1 - \delta \\
\implies Pr\Big(f(X) \leq Mf(X) + 2||X||t + 1\Big) &\geq 1 - \delta \\
\implies Pr\Big(f(X) - Mf(X) \leq 2||X||t + 1\Big) &\geq 1 - \delta
\end{aligned}
$$

Finally we rewrite the norm and insert the value we found for t so that

$$Pr\Big(f(X) - Mf(X) \leq \sqrt{16 \log \frac{2}{\delta}} \cdot \sqrt{\sum_{i=1}^{n} X_i^2} + 1\Big) \geq 1 - \delta$$

We compare the inequalities from both methods for the same δ

$$\text{Azuma's: } Pr\left(f(X) - Ef(X) \leq \sqrt{2 \log \frac{1}{\delta}} \cdot \sqrt{n}\right) \geq 1 - \delta \tag{2}$$

$$\text{Talagrand's: } Pr\Big(f(X) - Mf(X) \leq \sqrt{16 \log \frac{2}{\delta}} \cdot \sqrt{\sum_{i=1}^{n} X_i^2} + 1\Big) \geq 1 - \delta \tag{3}$$

In order to compare these statements we need to consider their differences. The +1 in (3) will not matter since one extra bin is the smallest possible measurement of difference. We are generally talking about many bins, so the number of bins will certainly deviate from the average by many as well. Also we can assume that the mean of f is close to the median. And since the logarithm dramatically shrinks its argument $\sqrt{2 \log \frac{1}{\delta}} \approx \sqrt{16 \log \frac{2}{\delta}}$.

We have concluded that the most significant difference between our concentration inequalities is $\sqrt{n}$ in Azuma and $\sqrt{\sum_{i=1}^{n} X_i^2}$ in Talagrand. Remember that X_i is distributed between 0 and 1 so at worst $\sqrt{\sum_{i=1}^{n} X_i^2} = \sqrt{\sum_{i=1}^{n} 1^2} = \sqrt{n}$. Therefore with Talagrand we already have a tighter bound with the same probability. But in general, X_i will not be close to one and X_i^2 will be even smaller.

Now consider the distribution with which X_i is chosen. It could produce most values extremely close to 0 and very few values close to one. This will put the constant $\sqrt{\sum_{i=1}^{n} X_i^2}$ close to zero and give us a very tight bound with Talagrand's inequality. Meanwhile the constant $\sqrt{n}$ from Azuma's inequality will not change at all as X_i changes. It cannot detect the distribution with which X_i is chosen. So we get the same bound whether the random values are all close to 1 or close to 0. Finally we can see the power of Talagrand's inequality in that it truly measures how each variable contributes to its expectation rather than treating each "dimension" the same.

5 Details

5.1 Hölder's Inequality

Theorem (Hölder's Inequality)**.** *Given a probability space Ω and measurable functions f and g if $p, q \in [1, \infty]$ satisfy $1/p + 1/q \leq 1$ then*

$$||fg||_1 \leq ||f||_p ||g||_q$$

When we make use of the above in the proof of theorem 2 our norm is the integral with respect to the product measure on OLD. We also chose $p = 1/(1-\lambda)$ and $q = 1/\lambda$ which are greater than one since $0 \leq \lambda \leq 1$ and $1/p + 1/q = (1-\lambda) + \lambda = 1 \leq 1$. So our proof above uses Hölder's Inequality in the following way

$$\begin{aligned}\int_{OLD} \exp(c_1)^{(1-\lambda)} \exp(c_2)^{\lambda} &\leq \left(\int_{OLD} \exp(c_1)^{\frac{1-\lambda}{1-\lambda}}\right)^{(1-\lambda)} \left(\int_{OLD} \exp(c_2)^{\frac{\lambda}{\lambda}}\right)^{\lambda} \\ &= \left(\int_{OLD} \exp(c_1)\right)^{(1-\lambda)} \left(\int_{OLD} \exp(c_2)\right)^{\lambda}\end{aligned}$$

5.2 Optimization

Lemma. $e^{(1-\lambda)^2/4}\gamma^{-\lambda} \leq 2 - \gamma$ *at the minimum over* λ

Proof. First we minimize with respect to λ

$$\begin{aligned}\frac{d}{dx} e^{(1-\lambda)^2/4}\gamma^{-\lambda} &= 0 \\ \implies \frac{1}{2} e^{(1-\lambda)^2/4}\gamma^{-\lambda}(-2\log\gamma + \lambda - 1) &= 0 \\ \implies \lambda &= 1 + 2\log\gamma\end{aligned}$$

But we also have the requirement that $0 \leq \lambda \leq 1 \implies e^{-1/2} \leq \gamma \leq 1$. If γ does not satisfy this condition then the minimum occurs at $\lambda = 0$.

If $\lambda = 0$ is the minimum then $0 \leq \gamma \leq e^{-1/2}$ and $e^{(1-\lambda)^2/4}\gamma^{-\lambda} = e^{1/4} \leq 2 - \gamma$.

If $\lambda = 1 + 2\log\gamma$ then

$$e^{(1-\lambda)^2/4}\gamma^{-\lambda} = e^{2\log\gamma}\gamma^{-1-2\log\gamma} = \gamma^2\gamma^{-1-2\log\gamma} = \gamma^{-2-4\log\gamma}$$

This has maximum value of $e^{1/4}$ at $\gamma = e^{-1/4}$. So for this value of λ we again have $e^{(1-\lambda)^2/4}\gamma^{-\lambda} \leq 2 - \gamma$ $\square$

5.3 Convex distance is Euclidean

The following is a discreet space where the convex distance is actually Euclidean and can reinforce our understanding of its behaviour. Consider $\Omega = \{0, 1\}^n$ and $A \subseteq \Omega$.

Consider the second version of the convex distance. We can think of our point x moving to the origin and points of A moving to the binary hypercube. Since A is discrete it will look almost the same in $U(A, x)$ just rotated. Except $U(A, x)$ could have some extra points with a 1 in some coordinates but conveniently these are further away from the origin and do not contribute to the convex distance. So our convex distance is the Euclidean distance from the origin to $V(A, x)$, the convex hull of $U(A, x)$. This is the exact same as the Euclidean distance from x to the convex hull of A.

References

[1] Alon, Noga and Spencer, Joel. *The Probabilistic Method*. John Wiley & Sons, Inc., 2008.

[2] Boucheron, Stéphane et al. *Concentration Inequalities: a Nonasymptotic Theory of Independence*. Oxford University Press, 2016.

[3] Cook, Nick. *Notes on Talagrand's Isoperimetric Inequality*. UCLA, 2017. `http://www.math.ucla.edu/~nickcook/talagrand.pdf`.

[4] Gamarnik, David. *Concentration Inequalities and Applications* . MIT, 2013. `https://ocw.mit.edu/courses/sloan-school-of-management/15-070j-advanced-stochastic-processes-fall-2013/lecture-notes/MIT15_070JF13_Lec13.pdf`.