1. COMPARISON INEQUALITIES

The study of the maximum (or supremum) of a collection of Gaussian random variables is of fundamental importance. In such cases, certain comparison inequalities are helpful in reducing the problem at hand to the same problem for a simpler correlation matrix. We start with a lemma of this kind and from which we derive two important results - Slepian's inequality and the Sudakov-Fernique inequality[8].

Lemma 1 (J.P. Kahane). *Let X and Y be $n \times 1$ mutivariate Gaussian vectors with equal means, i.e., $\mathbf{E}[X_i] = \mathbf{E}[Y_i]$ for all i. Let $A = \{(i,j) : \sigma^X_{ij} < \sigma^Y_{ij}\}$ and let $B = \{(i,j) : \sigma^X_{ij} > \sigma^Y_{ij}\}$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be any C^2 function all of whose partial derivatives up to second order have subgaussian growth and such that $\partial_i \partial_j f \geq 0$ for all $(i,j) \in A$ and $\partial_i \partial_j f \leq 0$ for all $(i,j) \in B$. Then, $\mathbf{E}[f(X)] \leq \mathbf{E}[f(Y)]$.*

Proof. First assume that both X and Y are centered. Without loss of generality we may assume that X and Y are defined on the same probability space and independent of each other.

Interpolate between them by setting $Z(\theta) = (\cos\theta)X + (\sin\theta)Y$ for $0 \leq \theta \leq \frac{\pi}{2}$ so that $Z(0) = X$ and $Z(\pi/2) = Y$. Then,

$$\mathbf{E}[f(Y)] - \mathbf{E}[f(X)] = \mathbf{E}\left[\int_0^{\pi/2} \frac{d}{d\theta} f(Z(\theta)) d\theta\right] = \int_0^{\pi/2} \frac{d}{d\theta} \mathbf{E}[f(Z_\theta)] d\theta.$$

The interchange of expectation and derivative etc., can be justified by the conditions on f but we shall omit these routine checks. Further,

$$\frac{d}{d\theta}\mathbf{E}[f(Z_\theta)] = \mathbf{E}[\nabla f(Z_\theta) \cdot \dot{Z}(\theta)] = \sum_{i=1}^n \left\{-(\sin\theta)\mathbf{E}[X_i \partial_i f(Z_\theta)] + (\cos\theta)\mathbf{E}[Y_i \partial_i f(Z_\theta)]\right\}.$$

Now use Exercise 14 to deduce (apply the exercise after conditioning on X or Y and using the independence of X and Y) that

$$\mathbf{E}[X_i \partial_i f(Z_\theta)] = (\cos\theta) \sum_{j=1}^n \sigma^X_{ij} \mathbf{E}[\partial_i \partial_j f(Z_\theta)]$$

$$\mathbf{E}[Y_i \partial_i f(Z_\theta)] = (\sin\theta) \sum_{j=1}^n \sigma^Y_{ij} \mathbf{E}[\partial_i \partial_j f(Z_\theta)].$$

Consequently,

$$\frac{d}{d\theta}\mathbf{E}[f(Z_\theta)] = (\cos\theta)(\sin\theta) \sum_{i,j=1}^n \mathbf{E}[\partial_i \partial_j f(Z_\theta)] \left(\sigma^Y_{ij} - \sigma^X_{ij}\right). \tag{1}$$

The assumptions on $\partial_i \partial_j f$ ensure that each term is non-negative. Integrating, we get $\mathbf{E}[f(X)] \leq \mathbf{E}[f(Y)]$.

It remains to consider the case when the means are not zero. Let $\mu_i = \mathbf{E}[X_i] = \mathbf{E}[Y_i]$ and set $\hat{X}_i = X_i - \mu_i$ and $\hat{Y}_i = Y_i - \mu_i$ and let $g(x_1, \ldots, x_n) = f(x_1 + \mu_1, \ldots, x_n + \mu_n)$. Then $f(X) = g(\hat{X})$ and $f(Y) = g(\hat{Y})$ while $\partial_i \partial_j g(x) = \partial_i \partial_j f(x + \mu)$. Thus, the already proved statement for centered variables implies the one for non-centered variables. ∎

Special cases of this lemma are very useful. We write X^* for $\max_i X_i$.

Corollary 2 (Slepian's inequality). *Let X and Y be $n \times 1$ mutivariate Gaussian vectors with equal means, i.e., $\mathbf{E}[X_i] = \mathbf{E}[Y_i]$ for all i. Assume that $\sigma^X_{ii} = \sigma^Y_{ii}$ for all i and that $\sigma^X_{ij} \geq \sigma^Y_{ij}$ for all i, j. Then,*

(1) *For any real $t_1, \ldots, t_n$, we have $\mathbf{P}\{X_i < t_i \text{ for all } i\} \geq \mathbf{P}\{Y_i < t_i \text{ for all } i\}$.*

(2) *$X^* \prec Y^*$, i.e., $\mathbf{P}\{X^* > t\} \leq \mathbf{P}\{Y^* > t\}$ for all t.*

---

[8] The presentation here is cooked up from Ledoux-Talagrand (the book titled *Probability on Banach spaces*) and from Sourav Chatterjee's paper on Sudakov-Fernique inequality. Chatterjee's proof can be used to prove Kahane's inequality too, and consequently Slepian's, and that is the way we present it here.

Proof. In the language of Lemma 1 by taking $B \subseteq \{(i,i) : 1 \le i \le n\}$while $A = \emptyset$. We would like to say that the first conclusion follows by simply taking $f(x_1,\dots,x_n) = \prod_{i=1}^n \mathbf{1}_{x_i<t_i}$. The only wrinkle is that it is not smooth. by approximating the indicator with smooth increasing functions, we can get the conclusion.

To elaborate, let $\psi \in \mathbb{C}^\infty(\mathbb{R})$ be an increasing function $\psi(t) = 0$ for $t < 0$ and $\psi(t) = 1$ for $t > 1$. Then $\psi_\varepsilon(t) = \psi(t/\varepsilon)$ increases to $\mathbf{1}_{t<0}$ as $\varepsilon \downarrow 0$. If $f_\varepsilon(x_1,\dots,x_n) = \prod_{i=1}^n \psi_\varepsilon(x_i - t_i)$, then $\partial_{ij} f \ge 0$ and hence Lemma 1 applies to show that $\mathbf{E}[f_\varepsilon(X)] \le \mathbf{E}[f_\varepsilon(Y)]$. Let $\varepsilon \downarrow 0$ and apply monotone convergence theorem to get the first conclusion.

Taking $t_i = t$, we immediately get the second conclusion from the first. ■

Here is a second corollary which generalizes Slepian's inequality (take $m = 1$).

Corollary 3 (Gordon's inequality). *Let $X_{i,j}$ and $Y_{i,j}$ be $m \times n$ arrays of joint Gaussians with equal means. Assume that*

1. *$Cov(X_{i,j}, X_{i,\ell}) \ge Cov(Y_{i,j}, Y_{i,\ell})$,*
2. *$Cov(X_{i,j}, X_{k,\ell}) \le Cov(Y_{i,j}, Y_{k,\ell})$ if $i \ne k$,*
3. *$Var(X_{i,j}) = Var(Y_{i,j})$.*

Then

1. *For any real $t_{i,j}$ we have $\mathbf{P}\left\{\bigcap_i \bigcup_j \{X_{i,j} < t_{i,j}\}\right\} \ge \mathbf{P}\left\{\bigcap_i \bigcup_j \{Y_{i,j} < t_{i,j}\}\right\}$,*
2. *$\min_i \max_j X_{i,j} \prec \min_i \max_j Y_{i,j}$.*

Exercise 4. Deduce this from Lemma 1.

Remark 5. The often repeated trick that we referred to is of constructing the two random vectors independently on the same space and interpolating between them. Then the comparison inequality reduces to a differential inequality which is simpler to deal with. Quite often different parameterizations of the same interpolation are used, for example $Z_t = \sqrt{1-t^2}X + tY$ for $0 \le t \le 1$ or $Z_s = \sqrt{1-e^{-2s}}X + e^{-s}Y$ for $-\infty \le s \le \infty$.

2. SUDAKOV-FERNIQUE INEQUALITY

Studying the maximum of a Gaussian process is a very important problem. Slepian's (or Gordon's) inequality helps to control the maximum of our process by that of a simpler process. For example, if $X_1,\dots,X_n$ are standard normal variables with positive correlation between any pair of them, then $\max X_i$ is stochastically smaller than the maximum of n independent standard normals (which is easy). However, the conditions of Slepian's inequality are sometimes restrictive, and the conclusions are much stronger than required. The following theorem is a more applicable substitute.

Theorem 6 (Sudakov-Fernique inequality). *Let X and Y be $n \times 1$ Gaussian vectors satisfying $\mathbf{E}[X_i] = \mathbf{E}[Y_i]$ for all i and $\mathbf{E}[(X_i - X_j)^2] \le \mathbf{E}[(Y_i - Y_j)^2]$ for all $i \ne j$. Then, $\mathbf{E}[X^*] \le \mathbf{E}[Y^*]$.*

Remark 7. Assume that the means are zero. If $\mathbf{E}[X_i^2] = \mathbf{E}[Y_i^2]$ for all i, then the condition $\mathbf{E}[(X_i - X_j)^2] \le \mathbf{E}[(Y_i - Y_j)^2]$ is the same as $\mathbf{E}[X_iX_j] \ge \mathbf{E}[Y_iY_j]$. Then Slepian's inequality would apply and we would get the much stronger conclusion of $X^* \prec Y^*$. The point here is the relaxing of the assumption of equal variances and settling for the weaker conclusion which only compares expectations of the maxima.

Proof. The proof of Lemma 1 can be copied exactly to get (1) for any smooth function f with appropriate growth conditions. Now we specialize to the function $f_\beta(x) = \frac{1}{\beta}\log\sum_{i=1}^n e^{\beta x_i}$ where $\beta > 0$ is fixed. Let $p_i(x) =$

$\frac{e^{\beta x_i}}{\sum_{i=1}^n e^{\beta x_i}}$, so that $(p_1(x),\dots,p_n(x))$ is a probability vector for each $x\in\mathbb{R}^n$. Observe that

$$\partial_i f(x) = p_i(x)$$
$$\partial_i\partial_j f(x) = \beta p_i(x)\delta_{i,j} - \beta p_i(x)p_j(x).$$

Thus, (1) gives

$$\frac{1}{\beta(\cos\theta)(\sin\theta)}\frac{d}{d\theta}\mathbf{E}[f_\beta(Z_\theta)] = \sum_{i,j=1}^n(\sigma_{ij}^Y-\sigma_{ij}^X)\mathbf{E}\left[p_i(x)\delta_{i,j}-p_i(x)p_j(x)\right]$$
$$= \sum_{i=1}^n(\sigma_{ii}^Y-\sigma_{ii}^X)\mathbf{E}[p_i(x)] - \sum_{i,j=1}^n(\sigma_{ij}^Y-\sigma_{ij}^X)\mathbf{E}[p_i(x)p_j(x)]$$

Since $\sum_i p_i(x)=1$, we can write $p_i(x)=\sum_j p_i(x)p_j(x)$ and hence

$$\frac{1}{\beta(\cos\theta)(\sin\theta)}\frac{d}{d\theta}\mathbf{E}[f_\beta(Z_\theta)] = \sum_{i,j=1}^n(\sigma_{ii}^Y-\sigma_{ii}^X)\mathbf{E}[p_i(x)p_j(x)] - \sum_{i,j=1}^n(\sigma_{ij}^Y-\sigma_{ij}^X)\mathbf{E}[p_i(x)p_j(x)]$$
$$= \sum_{i<j}\mathbf{E}[p_i(x)p_j(x)]\left(\sigma_{ii}^Y-\sigma_{ii}^X+\sigma_{jj}^Y-\sigma_{jj}^X-2\sigma_{ij}^Y+2\sigma_{ij}^X\right)$$
$$= \sum_{i<j}\mathbf{E}[p_i(x)p_j(x)]\left(\gamma_{ij}^X-\gamma_{ij}^Y\right)$$

where $\gamma_{ij}^X=\sigma_{ii}^X+\sigma_{jj}^X-2\sigma_{ij}^X=\mathbf{E}[(X_i-\mu_i-X_j+\mu_j)^2]$. Of course, the latter is equal to $\mathbf{E}[(X_i-X_j)^2]-(\mu_i-\mu_j)^2$. Since the μ_i are the same for X as for Y we get $\gamma_{ij}^X\le\gamma_{ij}^Y$. Clearly $p_i(x)\ge 0$ too. Therefore, $\frac{d}{d\theta}\mathbf{E}[f_\beta(Z_\theta)]\ge 0$ and we get $\mathbf{E}[f_\beta(X)]\le\mathbf{E}[f_\beta(Y)]$. Letting $\beta\uparrow\infty$ we get $\mathbf{E}[X^*]\le\mathbf{E}[Y^*]$. ■

Remark 8. This proof contains another useful idea - to express $\max_i x_i$ in terms of $f_\beta(x)$. The advantage is that f_β is smooth while the maximum is not. And for large β, the two are close because $\max_i x_i \le f_\beta(x)\le \max_i x_i+\frac{\log n}{\beta}$.

If Sudakov-Fernique inequality is considered a modification of Slepian's inequality, the analogous modification of Gordon's inequality is the following. We leave it as exercise as we may not use it in the course.

Exercise 9. (optional) Let $X_{i,j}$ and $Y_{i,j}$ be $n\times m$ arrays of joint Gaussians with equal means. Assume that

(1) $\mathbf{E}[|X_{i,j}-X_{i,\ell}|^2]\ge\mathbf{E}[|Y_{i,j}-Y_{i,\ell}|^2]$,

(2) $\mathbf{E}[|X_{i,j}-X_{k,\ell}|^2]\le\mathbf{E}[|Y_{i,j}-Y_{k,\ell}|^2]$ if $i\ne k$.

Then $\mathbf{E}[\min_i\max_j X_{i,j}]\le\mathbf{E}[\min_i\max_j Y_{i,j}]$.