STOR 655 Homework 2022

A. Calculus and Real Analysis

1. Show that $1+x \leq e^x$ for every real number x. First sketch the picture. Then use calculus to rigorously establish the result. Deduce that $\log x \leq x - 1$ for every $x > 0$.

2. Show that $(1 + u/3)^3 \geq 1 + u$ for every $u \geq 0$.

3. Recall that if $f : \mathcal{X} \to \mathbb{R}$ is a real-valued function then the argmax of f is the set of points in x at which f is maximized,

$$\arg\max_{x \in \mathcal{X}} f(x) = \left\{ x \in \mathcal{X} : f(x) = \sup_{u \in \mathcal{X}} f(u) \right\}.$$

The argmin of f is similarly defined.

(a) Let $f : \mathcal{X} \to \mathbb{R}$ be defined on a set $\mathcal{X} \subseteq \mathbb{R}$ by $f(x) = x^2$. Identify the value of

$$\sup_{x \in \mathcal{X}} f(x) \quad \text{and} \quad \arg\max_{x \in \mathcal{X}} f(x)$$

in each of the following cases: $\mathcal{X} = [-2, 2]$, $\mathcal{X} = (-2, 2]$, $\mathcal{X} = (-2, 2)$, and $\mathcal{X} = (-3, 2]$.

(b) Let A be a bounded subset of $\mathbb{R}^d$. Identify the values of $\inf_x f(x), \sup_x f(x), \arg\min_x f(x)$, and $\arg\max_x f(x)$ for the function $f : \mathbb{R}^d \to \mathbb{R}$ defined by

$$f(x) = \inf_{y \in A} ||x - y||.$$

4. How you can obtain inequalities for $\log(1 + x)$ and $\log(1 - x)$ from Taylor's theorem.

(a) Expand the function $h(v) = \log v$ in a third order Taylor series around the point $v = 1$. (Thus you will be expressing $h(1 + x)$ in terms of x, $h(1)$, $h^{'}(1)$, $h^{''}(1)$, and $h^{'''}(u)$ for some u between 1 and $1 + x$. Note that x may be negative.)

(b) By examining the final term in the series, use part (a) to show that $\log(1+x) \geq x - x^2/2$ for $x \geq 0$.

(c) By examining the final term in the series, use part (a) to show that $\log(1 - x) \leq -x - x^2/2$ for $0 \leq x < 1$.

5. Let $h(u) = (1+u)\log(1+u) - u$. (This function appears in Bennett's exponential inequality for sums of independent, bounded random variables.)

(a) By considering the first few terms of Taylor expansion of the function $h(\cdot)$ around zero, show that for every $u \geq 0$

$$h(u) \geq \frac{u^2}{2+2u}$$

(b) (Optional) Use calculus to establish the stronger bound that for every $u \geq 0$

$$h(u) \geq \frac{u^2}{2+2u/3}$$

6. Show that $xy \leq 3x^2 + y^2/3$ for $x, y \geq 0$.

7. Show that if $d \geq 3$ then $\int_{\mathbb{R}^d} ||u||^{-2}\, e^{-||u||^2} du = c \int_0^\infty e^{-r^2} r^{d-3} dr < \infty$.

8. Show that $|e^a - e^b| \leq e^b\, e^{|a-b|}\, |a-b|$.

9. Let $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ be two sequences of numbers. Rigorously establish the following inequalities.

(a) $\min\{a_i\} + \min\{b_i\} \;\leq\; \min\{a_i + b_i\} \;\leq\; \min\{a_i\} + \max\{b_i\}$

(b) $-\min\{a_i\} = \max\{-a_i\}$ and $-\max\{a_i\} = \min\{-a_i\}$

(c) $\max\{a_i\} - \max\{b_i\} \;\leq\; \max\{|a_i - b_i|\}$

Use part (b) to find a chain of inequalities like that in part (a) for maxima

10. (Saddle points and minimax) Let $\mathcal{X}$ and $\mathcal{Y}$ be sets and let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be any function.

a. Show that, with no further assumptions,

$$\sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} f(x,y) \;\leq\; \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} f(x,y) \tag{1}$$

This simple fact plays an important role in optimization, where it implies the weak duality property of the Lagrange dual problem, and in game theory, where it has connections with Nash equilibria. A pair $(\tilde{x}, \tilde{y}) \in \mathcal{X} \times \mathcal{Y}$ is called a *saddle point* for f if

$$f(\tilde{x}, y) \;\leq\; f(\tilde{x}, \tilde{y}) \;\leq\; f(x, \tilde{y}) \quad \text{for every } x \in \mathcal{X} \text{ and } y \in \mathcal{Y}$$

b. Show that if $(\tilde{x}, \tilde{y})$ is a saddle point for f then

$$f(\tilde{x}, \tilde{y}) \;=\; \inf_{x \in \mathcal{X}} f(x, \tilde{y}) \quad \text{and} \quad f(\tilde{x}, \tilde{y}) \;=\; \sup_{y \in \mathcal{Y}} f(\tilde{x}, y)$$

To see how these inequalities explain the use of the terminology "saddle point", assume that f is nice and smooth, and sketch what it will look like in a neighborhood around the point $(\tilde{x}, \tilde{y})$.

c. Show that the existence of a saddle point implies equality in inequality (1) above.

d. Evaluate both sides of (1) when $\mathcal{X} = [0, 1]$, $\mathcal{Y} = [-1, 1]$, and $f(x, y) = x^2 y$.

11. For each $k = 1, \ldots, K$ let $\{a_k(n) : n \geq 1\}$ be a sequence of real numbers. Find an inequality or equality relating

$$\limsup_{n \to \infty} \max_{1 \leq k \leq K} a_k(n) \quad \text{and} \quad \max_{1 \leq k \leq K} \limsup_{n \to \infty} a_k(n)$$

Find an inequality or equality relating

$$\liminf_{n \to \infty} \max_{1 \leq k \leq K} a_k(n) \quad \text{and} \quad \max_{1 \leq k \leq K} \liminf_{n \to \infty} a_k(n)$$

12. Let $a_1, \ldots, a_n$ be real numbers. Show that $n^{-1} \sum_{k=1}^{n} |a_k| \leq (n^{-1} \sum_{k=1}^{n} a_k^2)^{1/2}$.

13. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be positive constants.

a. Use Jensen's inequality to establish the Arithmetic-Geometric mean inequality

$$\frac{1}{n} \sum_{i=1}^{n} a_i \;\geq\; \left(\prod_{i=1}^{n} a_i \right)^{1/n} .$$

b. Establish the inequality

$$(\Pi_{k=1}^{n} a_k)^{1/n} \;+\; (\Pi_{k=1}^{n} b_k)^{1/n} \;\leq\; (\Pi_{k=1}^{n} (a_k + b_k))^{1/n}$$

Hint: First divide the LHS by the RHS.

14. Let $||u|| = \langle u, u \rangle^{1/2}$ be the usual Euclidean norm on $\mathbb{R}^d$. Establish the following.

(a) $||u|| \geq 0$ with equality iff $u = 0$

(b) $||u + v||^2 = ||u||^2 + 2\langle u, v \rangle + ||v||^2$

(c) Cauchy-Schwartz inequality $|\langle u, v\rangle| = |u^t v| \leq ||u||\,||v||$

(d) $||u+v|| \leq ||u|| + ||v||$ Hint: square the left side and use Cauchy-Schwartz

(e) $|\,||u|| - ||v||\,| \leq ||u - v||$ (reverse triangle inequality)

15. Let $x = (x_1, \ldots, x_d)^t \in \mathbb{R}^d$ and let $||x||$ be the Euclidean (ℓ_2) norm of x. Show that for $1 \leq i \leq d$,

$$|x_i| \;\leq\; ||x|| \;\leq\; |x_1| + \cdots + |x_d|.$$

Use the inequalities to show that if $X \in \mathbb{R}^d$ is a random vector then $\mathbb{E}||X|| < \infty$ if and only if $\mathbb{E}|X_i| < \infty$ for $1 \leq i \leq d$.

16. Show that $||x||_\infty = \lim_{p \nearrow \infty} ||x||_p$. For $0 \leq p \leq 1$ define $||x|_p = \sum_{i=1}^d |x_i|^p$. Show that $||x||_0 = \lim_{p \searrow 0} ||x||_p$.

17. Establish the following linear algebra facts from class. Let $A, B \in \mathbb{R}^{n \times n}$.

(a) If A is a projection matrix then all of its eigenvalues are zero or one.

(b) If A is a projection matrix then $\text{rank}(A) = \text{tr}(A)$.

(c) If A is a symmetric projection matrix then $A\,v$ is orthogonal to $v - A\,v$ for every v.

(d) $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$

(e) $\text{tr}(A\,B) = \text{tr}(B\,A)$

(f) If $A > 0$ then $A^{-1} > 0$.

18. Show that if $Q \in \mathbb{R}^{n \times n}$ is orthogonal then $||Qx|| = ||x||$ for every x. What does this tell you about the real eigenvalues of Q? Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Use the spectral decomposition of A to show that

$$\sup_{x:\, x^T x = 1} x^T A x \;=\; \lambda_n$$

where λ_n is the largest eigenvalue of A. Deduce from this that

$$\sup_{x \neq 0} \frac{x^T A x}{x^T x} \;=\; \lambda_n.$$

Find a vector for which the inequality is satisfied with equality.

19. Show that if $u, v \in \mathbb{R}^n$ are orthogonal then $||u||_2 + ||v||_2 \leq \sqrt{2}||u + v||_2$.

20. Let $A(t) = \{A_{i,j}(t) : 1 \leq i, j \leq n\}$ be a matrix whose entries are differentiable functions of a real number t. Define the entry-wise derivative

$$A'(t) \;=\; \{A'_{i,j}(t) : 1 \leq i, j \leq n\}$$

(a) Show that the entry-wise derivate obeys the usual product rule, that is,

$$[A(t)B(t)]' \;=\; A(t)B'(t) \,+\, A'(t)B(t)$$

21. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function. Give the definition of what it means for f to be (i) continous and (ii) uniformly continuous.

22. Give a simple example of a family of functions $g_n : \mathbb{R} \to [0, 1]$ such that $g_n(x) \to g(x) = 1$ for each $x \in \mathbb{R}$ but $\sup_{x \in \mathbb{R}} |g_n(x) - g(x)| = 1$ for each n. *Optional:* Find an example like that above with functions $g_n : [0, 1] \to [0, 1]$.

23. Let $A \subset \mathbb{R}^d$ be non-empty. Define the function $f : \mathbb{R}^d \to [0, \infty)$, representing the minimum distance from x to the set A, by

$$f(x) := \inf_{y \in A} ||x - y||$$

Show that $f(x)$ is Lipschitz with constant 1, that is, $|f(x) - f(y)| \leq ||x - y||$ for every $x, y \in \mathbb{R}^d$.

24. Let $\mathcal{X} \subseteq \mathbb{R}^d$ and let $f_1, f_2, \ldots, f : \mathcal{X} \to \mathbb{R}$. Suppose that f_n converges uniformly to f in the sense that

$$\sup_{x \in \mathcal{X}} |f_n(x) - f(x)| \;\to\; 0 \;\text{ as }\; n \to \infty.$$

For each $n \geq 1$ let $x_n \in \arg\max f_n$, which we assume to be non-empty.

(a) Show that $\sup_x f(x)$ is finite and that $f(x_n) \to \sup_{x \in \mathcal{X}} f(x)$.

(b) Show that if f is continuous and $\mathcal{X}$ is compact, then $d(x_n, \arg\max f) \to 0$, where $d(x, K) = \inf_{u \in K} ||x - u||$.

25. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function. Show that the following are equivalent.

a f is upper semicontinuous (as defined in the lecture notes) on $\mathbb{R}^d$

b for every $x_0 \in \mathbb{R}^d$ and every $\epsilon > 0$ there is a $\delta > 0$, possibly depending on x, such that $||x - x_0|| < \delta$ implies $f(x) \leq f(x_0) + \epsilon$

c the super-level sets $\{x : f(x) \geq \alpha\}$ are closed for every $\alpha \in \mathbb{R}$

26. Show that if $f_1, f_2, \ldots : \mathbb{R}^d \to \mathbb{R}$ are u.s.c. then so is $g(x) = \inf_n f_n(x)$.

27. Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is upper semicontinuous at v. Show that $\sup_{u \in B(v,\delta)} f(u) \searrow f(v)$ as $\delta \to 0$, where $B(v, \delta)$ is the open ball of radius δ centered at v.

28. Let (S, d) be a metric space, and let $N(S, \epsilon)$ be the covering number of S under the metric $d(., .)$ at radius ϵ.

(a) What can you say about the limit of $N(S, \epsilon)$ as $\epsilon \to 0$? [Consider the case where S is finite and S is infinite.]

(b) Now let $S_0 \subseteq S$ be a subset of S. By definition, an ϵ-cover of S_0 contains of balls of radius ϵ centered at points in S_0, and $N(S_0, \epsilon)$ is the size of the smallest such cover. Consider instead general ϵ-covers of S_0 that are centered at points in S, so that centers need *not* be in S_0. Let $\tilde{N}(S_0, \epsilon)$ be the smallest such general cover. Find a simple relationship between $N(S_0, \epsilon)$ and $\tilde{N}(S_0, \epsilon)$.

29. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be numbers in the interval $[-1, 1]$. Establish the inequality

$$|a_1 \cdots a_n - b_1 \cdots b_n| \;\leq\; \sum_{i=1}^{n} |a_i - b_i|$$

Hint: Use induction and the fact that $a_1 a_2 - b_1 b_2 = (a_1 - b_1)a_2 + b_1(a_2 - b_2)$.

B. The Normal Distribution

1. Let $U \sim \mathcal{N}_d(\mu, \Sigma)$ and let $V = \Sigma^{1/2} Y + \mu$ where $Y \sim \mathcal{N}_d(0, I)$.

(a) Show that $\mathbb{E}U = \mathbb{E}V$ and that $\mathrm{Var}(U) = \mathrm{Var}(V)$.

(b) Fix $v \in \mathbb{R}^d$. Find the distributions of the random variables $\langle v, U\rangle$ and $\langle v, V\rangle$. Note that these distributions are the same. Thus $U \stackrel{\mathrm{d}}{=} V$.

2. Give a simple example of random vectors $X, Y \in \mathbb{R}^2$ such that $\text{Cov}(X, Y) \neq \text{Cov}(Y, X)$.

3. Let $X \sim \mathcal{N}(0, \sigma^2)$. Establish the identity

$$\mathbb{E} \exp\{aX^2 + bX\} = \frac{1}{\sqrt{1 - 2a\sigma^2}} \exp\left\{\frac{\sigma^2 b^2}{2(1 - 2a\sigma^2)}\right\}$$

Hint: Write the expectation as an integral. Combine terms in the exponent and complete the square. Remove the constant factor and perform a simple change of variables to evaluate the remaining integral.

4. Let $X \sim \mathcal{N}_d(\mu, \Sigma)$, and let $A \in \mathbb{R}^{k \times d}$ and $B \in \mathbb{R}^{l \times d}$ be matrices. Show that the random vectors $Y = AX$ and $Z = BX$ are independent if and only if $A \Sigma B^T = 0$. You may appeal to the general independence result from class.

5. Show that if $X \sim \mathcal{N}_d(\mu, \Sigma)$ and $U = X^T A X$ then $\mathbb{E} U = \text{tr}(A\Sigma) + \mu^T A \mu$. (It may be helpful to use the fact that $\text{tr}(UV) = \text{tr}(VU)$.)

6. (Stein's Lemma) Let $X \sim \mathcal{N}(0, 1)$ and let f be a continuously differentiable real-valued function such that $\mathbb{E}|f'(X)| < \infty$.

(a) Assuming that f is zero outside a finite interval (a, b), use integration-by-parts to establish that $\mathbb{E}[X f(X)] = \mathbb{E} f'(X)$.

(b) Extend the identity above to the case $X \sim \mathcal{N}(0, \sigma^2)$

(c) Show that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then $\mathbb{E}[(X - \mu) f(X)] = \sigma^2 \, \mathbb{E} f'(X)$

7. (Stein's Lemma for Covariance) Let $X, Y \in \mathbb{R}$ be non-degenerate jointly normal random variables with mean zero, and let f be a continuously differentiable real-valued function satisfying appropriate integrability conditions.

a. Argue that we can write $X = aZ_1 + bZ_2$ and $Y = bZ_1 + cZ_2$ where Z_1, Z_2 are independent standard normal random variables, and a, b, c are real constants.

b. Find $\text{Cov}(X, Y)$ in terms of a, b, c.

c. Show that $\text{Cov}(f(X), Y) = \mathbb{E} f'(X) \, \text{Cov}(X, Y)$. Hint: Use the representations of X and Y in terms of Z_1 and Z_2. Apply Stein's identity after appropriate conditioning.

d. Give some thought to what integrability conditions are needed for the covariance identity in part c.

8. (Bivariate normal distribution). Let $X = (X_1, X_2)^t \sim \mathcal{N}_2$ with

$$\mathbb{E}X_1 = \mu_1,\ \mathbb{E}X_2 = \mu_2,\ \mathrm{Var}(X_1) = \sigma_1^2,\ \mathrm{Var}(X_2) = \sigma_2^2,\ \mathrm{Corr}(X_1, X_2) = \rho \in [-1, 1]$$

(a) Find $\mu = \mathbb{E}X$ and $\Sigma = \mathrm{Var}(X)$ in terms of the quantities above.

(b) Find the determinant of Σ and conclude that Σ is invertible if and only if $\rho \in (-1, 1)$.

(c) Find Σ^{-1} when $\rho \in (-1, 1)$.

(d) Write down the density $f(x)$ of X in the case $\rho \in (-1, 1)$.

9. Let $\Phi(x)$ and $\phi(x)$ be the cumulative distribution function and density, respectively, of the standard normal distribution. In this problem, you are asked to find a useful approximation to $1 - \Phi(x)$ when x is large. Note that for $x > 0$,

$$1 - \Phi(x) \ = \ \Phi(-x) \ = \ \int_{-\infty}^{-x} \frac{1}{t} \cdot t\,\phi(t)\,dt$$

(a) Apply integration-by-parts to the last integral above. Use the resulting expression establish the upper bound $1 - \Phi(x) \leq x^{-1}\,\phi(x)$ for $x > 0$.

(b) Apply the same steps to the integral appearing in the integration-by-parts. Use this to establish the lower bound

$$1 - \Phi(x) \geq (\frac{1}{x} - \frac{1}{x^3})\,\phi(x) \quad \text{for} \quad x > 0.$$

(c) Conclude that as $x \to \infty$ $(1 - \Phi(x)) = \frac{\phi(x)}{x}(1 + o(1))$

10. Let $\Gamma(x)$ be the standard Gamma function, defined for $x > 0$. Show that if $Z \sim \mathcal{N}(0, 1)$ then for each $p \geq 1$

$$\mathbb{E}|Z|^p \ = \ \frac{2^{p/2}}{\sqrt{\pi}}\Gamma((1 + p)/2)$$

Deduce from this fact and Stirling's approximation that $||Z||_p := (\mathbb{E}|Z|^p)^{1/p} = O(p^{1/2})$.

11. Let $X_1, \ldots, X_n$ be independent standard normal random variables. Here we identify upper and lower bounds for the expectation of $K_n := \max_{1 \leq i \leq n} |X_i|$.

(a) Using the bound from class and the fact that $K_n = \max_i(X_i, -X_i)$ show that $\mathbb{E}K_n \leq (2\,\log 2n)^{1/2}$.

(b) Let $\Phi()$ be the CDF of the standard normal. Show that

$$K_n \;=\; \Phi^{-1}\left(\frac{1}{2}+\frac{1}{2}\max_{1\le i\le n} V_i\right)$$

where $V_1,\ldots,V_n$ are independent Uniform$(0,1)$ random variables.

(c) Show that $\Phi^{-1}(u)$ is convex on $[1/2,1)$. Apply Jensen's inequality to the expression in (b) to obtain the bound $\mathbb{E}K_n \ge \Phi^{-1}\left(1-1/(2n+2)\right)$.

(d) Show that $\Phi^{-1}(1-t^{-1})/(2\,\log t)^{1/2} \to 1$ as $t\to\infty$.

(e) Conclude from (a), (c), and (d) that $\mathbb{E}K_n/(2\,\log n)^{1/2} \to 1$ as $n\to\infty$.

12. *Extreme value theory for the Gaussian.* Let a_n and b_n be the extreme value scaling and centering constants for the maximum M_n of n independent standard Gaussian random variables.

(a) Fix $x\in\mathbb{R}$ and let $x_n = x/a_n + b_n$. Show that $n\,\phi(x_n)/x_n \to e^{-x}$ as n tends to infinity. [In your calculations, identify and pay careful attention to the leading order terms.]

(b) Using the result of part (a) and the standard Gaussian tail bound from an earlier homework, show that $n(1-\Phi(x_n)) \to e^{-x}$.

(c) Use part (b) and the lemma from lecture to show that as n tends to infinity

$$\mathbb{P}(a_n(M_n-b_n)\le x) \to G(x) = e^{-e^{-x}}$$

(d) Show that $G(x)$ is the CDF of $-\log V$ where $V\sim \mathrm{Exp}(1)$.

13. Establish the following facts about the Gaussian mean width $w(K)$ of a bounded set $K\subseteq\mathbb{R}^n$.

(a) If $K_1\subseteq K_2$ then $w(K_1)\le w(K_2)$

(b) $w(K)\ge 0$

(c) If $A\in\mathbb{R}^{n\times n}$ is orthogonal then $w(AK)=w(K)$

(d) For each $u\in\mathbb{R}^n$, $w(K+u)=w(K)$

(e) $w(K)=w(\mathrm{conv}(K))$

(f) $\sqrt{2/\pi}\,\mathrm{diam}(K)\le w(K)\le n^{1/2}\,\mathrm{diam}(K)$

(g) $w(K)\le 2\,\mathbb{E}\sup_{x\in K}\langle x,V\rangle$ with $V\sim\mathcal{N}_n(0,I)$

14. Read the statement and proof of the basic Gaussian comparison lemma in the on-line notes. Fill in the necessary details for equation (1.3), which makes use of Gaussian integration-by-parts. Write out a proof of the Gaussian comparison lemma in the case $n = 1$, following the proof of the general result. As $n = 1$, you will not need the conditioning argument, but you will need to exchange the operations of expectation and differentiation. Provide sufficient conditions on G and its derivatives to justify this exchange of limit operations, and show as carefully as you can why these conditions are sufficient. (You need not worry about finding the most general sufficient conditions; any reasonable conditions will do.)

15. Carefully verify that the Gaussian comparison lemma holds for the quadratic function $G(x) = x^t A x$, where A is a symmetric matrix.

16. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \sim \mathcal{N}(\theta_i, 1)$. Suppose that we wish to simultaneously test the hypotheses $\mathrm{H}_{0,i} : \theta_i = 0$ vs. $\mathrm{H}_{1,i} : \theta_i \neq 0$ for $1 \leq i \leq n$. Consider a simple threshold test in which we reject $\mathrm{H}_{0,i}$ if $|X_i| > \tau$ and accept $\mathrm{H}_{0,i}$ otherwise. Using the asymptotic results on Gaussian extreme values, find a value of the threshold τ, depending on n, so that the family-wise error rate of the test under the global null $\theta_1 = \cdots = \theta_n = 0$ is (approximately) controlled at 5%.

17. Use the general versions of Stein's Lemma given in class to show that if $Y \sim \mathcal{N}_n(\theta, I)$ and $g : \mathbb{R}^n \to \mathbb{R}^n$ is a sufficiently nice function then $\mathbb{E}[(Y - \theta)^T g(Y)] = \mathbb{E}[\nabla^T g(Y)]$.

18. In class we established a risk bound for the James-Stein estimator for observations $Y \sim \mathcal{N}_n(\theta, I)$. By looking over the proof, establish an analogous bound in the case $Y \sim \mathcal{N}_n(\theta, \sigma^2 I)$ with $\sigma > 0$ known.

19. Show that if $Y \sim \mathcal{N}(0, \sigma^2)$ and $c > 0$ then $\mathbb{E}\{|Y| I(|Y| > c)\} \leq \sigma \exp\{-c^2/2\sigma^2\}$

C. Convex Sets and Functions

1. Let $\{C_\lambda : \lambda \in \Lambda\}$ be convex sets. Show that the intersection $C = \cap_{\lambda \in \Lambda} C_\lambda$ is convex.

2. Show that the following subsets of $\mathbb{R}^d$ are convex.

a. The emptyset

b. The hyperplane $H = \{x : x^t u = b\}$

c. The halfspace $H_+ = \{x : x^t u > b\}$

d. The ball $B(x_0, r) = \{x : ||x - x_0|| \leq r\}$

3. Show that if $f_1, \ldots, f_k$ are convex functions defined on the same set, and $w_1, \ldots, w_k$ are non-negative, then $f = \sum_{j=1}^k w_j \, f_j$ is convex.

4. Let $\{f_\lambda : \lambda \in \Lambda\}$ be convex functions defined on a common set C. Show that the supremum $f = \sup_{\lambda \in \Lambda} f_\lambda$ is convex.

5. Recall that the convex hull of a set $A \subseteq \mathbb{R}^d$, denoted $\text{conv}(A)$, is the intersection of all convex sets C containing A. Show that $\text{conv}(A)$ is equal to the set of all convex combinations $\sum_{i=1}^k \alpha_i \, x_i$, where $k \geq 1$ is finite, $x_1, \ldots, x_k \in A$, and the coefficients α_i are non-negative and sum to one.

6. Identify the extreme points (if any) of the following convex sets.

a. The hyperplane $H = \{x : x^t u = b\}$

b. The halfspace $H_+ = \{x : x^t u > b\}$

c. The closed ball $\overline{B}(x_0, r) = \{x : ||x - x_0|| \leq r\}$

7. Let $f : C \to \mathbb{R}$ be a strictly convex function defined on a convex set $C \subseteq \mathbb{R}^n$. Show that $\text{argmax}_{x \in C} f(x)$ is contained in the set of extreme points of C.

8. (Set sums and scaler products) Given sets $A, B \subseteq \mathbb{R}^d$ and a constant $\alpha \in \mathbb{R}$ define the set sum and set scaler product as follows:

$$A + B \;=\; \{x + y : x \in A \text{ and } y \in B\} \qquad \alpha A \;=\; \{\alpha x : x \in A\}$$

a. (Optional) Show that if A is open then $A + B$ is open regardless of whether B is open.

b. Show that if A and B are convex, then so is $A + B$.

c. If A is convex is $A + B$ necessarily convex?

d. Show by example that, in general, $2A \neq A + A$.

d. Show that if A is convex then $\alpha A + \beta A = (\alpha + \beta)A$ for all $\alpha, \beta \geq 0$.

9. Let f be a convex function defined on a convex set C. Show that for each $\alpha \in \mathbb{R}$ the level set $L(\alpha) = \{x : f(x) \leq \alpha\}$ is convex.

10. Let $X \in \mathbb{R}$ be an integrable random variable with CDF $F(x)$, and for $0 < p < 1$ let $h_p(x, \theta) = p(x - \theta)_+ + (1 - p)(\theta - x)_+$.

(a) Show that for each fixed p and x, $h_p(x, \theta)$ is a convex function of θ.

(b) Show that, under reasonable assumptions on F, the quantity $\mathbb{E}h_p(X, \theta)$ is minimized by the pth quantile $F^{-1}(p)$ of X. Clearly state any assumptions that you make.

(c) What does the result of part (b) tell you in the special case $p = 1/2$.

11. Let f be a convex function on an open interval $I \subseteq \mathbb{R}$ and let $a < b < c$ be in I.

(a) Show that

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(c) - f(a)}{c - a} \leq \frac{f(c) - f(b)}{c - b}.$$

(Hint: express b as a convex combination of a and c and then apply the definition of convexity.)

(b) Draw a picture illustrating this result. Interpret the result in terms of the slopes of chords of the function f.

(c) Let $L^*(b) = \sup_{a<b} L(a, b)$ and $U^*(b) = \inf_{c>b} U(c, b)$. Using equation (**??**) above, argue carefully that $L^*(b) \leq U^*(b)$ and that both quantities are finite.

(d) Argue that for every $c \in I$ with $c > b$ the inequality $f(c) \geq f(b) + (c - b)\, L^*(b)$ holds. Argue that for every $a \in I$ with $a < b$ the inequality $f(a) \geq f(b) + (a - b)\, U^*(b)$ holds.

D. Statistics

1. Establish the following relations for random vectors X and Y of appropriate dimension.

(a) $\mathbb{E}(AX) = A\mathbb{E}X$

(b) $\mathrm{Var}(AX) = A\,\mathrm{Var}(X)A^t$

(c) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)^t$

(d) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, X)$

(e) If X, Y are independent, then $\mathrm{Cov}(X, Y) = 0$

2. Let X, Y be non-negative random variables defined on the same probability space.

(a) Show that $\mathbb{E}X = \int_0^\infty \mathbb{P}(X > t)\, dt$. Hint: Use the identity $x = \int_0^\infty \mathbb{I}(x > t)\, dt$ in the integral for $\mathbb{E}X$.

(b) Let $g : [0, \infty) \to \mathbb{R}$ be a function with $g(0) = 0$ having a continuous, non-negative derivative $g'(x)$. Argue that $g(x)$ is non-negative and use the proof from part (a) to show that $\mathbb{E}g(X) = \int_0^\infty \mathbb{P}(X > t)\, g'(t)\, dt$

(c) (*Optional.*) Show that $\mathrm{Cov}(g(X), g(Y)) = \int_0^\infty \int_0^\infty H(s, t)\, g'(s)\, g'(t)\, ds\, dt$ where

$$H(s, t) = \mathbb{P}(X > s, Y > t) - \mathbb{P}(X > s)\, \mathbb{P}(Y > t)$$

3. Let U, V, W be random variables. Carefully establish the following inequalities.

(a) $\mathbb{P}(|U + V| > a + b) \;\leq\; \mathbb{P}(|U| > a) \;+\; \mathbb{P}(|V| > b)$ for every $a, b \geq 0$.

(b) $\mathbb{P}(|U\, V| > a) \;\leq\; \mathbb{P}(|U| > a/b) \;+\; \mathbb{P}(|V| > b)$ for every $a, b > 0$.

4. Let $X_1, X_2, \ldots, X$ and $Y_1, Y_2, \ldots, Y$ be d-dimensional random vectors defined on the same probability space such that $X_n \to X$ in probability and $Y_n \to Y$ in probability. Show that $(X_n + Y_n) \to (X + Y)$ in probability.

5. Let $F_1, F_2, \ldots, F$ be one dimensional CDFs. Show that if $F(x)$ is continuous, and $F_n(x) \to F(x)$ as n tends to infinity for every $x \in \mathbb{R}$, then $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0$ as n tends to infinity. [Hint: Mimic the arguments for the Glivenko-Cantelli theorem given in class.] What are the implications of this fact for the central limit theorem?

6. Establish the Glivenko-Cantelli theorem for an i.i.d. sequence $X_1, X_2, \ldots$ of discrete random variables taking values in a countable set $S \subseteq \mathbb{R}$. [Hint: The case when S is finite can be handled by a direct appeal to the LLN. If S is infinite, split S into a finite set S_0 and an infinite set S_1 with probability at most ϵ. Apply the LLN to handle S_0, and argue that any residual error arising from S_1 is comparable to ϵ.]

7. Let X be a real-valued random variable with CDF $F(x)$. For $0 < p < 1$ define the quantile function

$$\varphi(p) \;=\; \inf\{x : F(x) \geq p\}$$

(a) Use the right-continuity of F to show that $\varphi(p) \leq x$ if and only if $p \leq F(x)$.

A number $M = M(X)$ is said to be a median of X if $P(X > M) \leq 1/2$ and $P(X < M) \leq 1/2$. Note that X may have more than one median.

(b) Show that $M = M(X)$ always exists and that $M(X)$ is unique if F is monotone increasing.

8. Establish the following relations for stochastic order symbols

(a) $o_p(1) = O_p(1)$

(b) $O_p(1) + O_p(1) = O_p(1)$

(c) $O_p(1) + o_p(1) = O_p(1)$

(d) $o_p(1) + o_p(1) = o_p(1)$

(e) $O_p(1)O_p(1) = O_p(1)$

(f) $O_p(1)o_p(1) = o_p(1)$

9. Show directly (without appealing to results about weak convergence) that if $X_1, X_2, \ldots, X \in \mathbb{R}^d$ are random vectors such that $X_n \to X$ in probability then $X_n = O_p(1)$.

10. Let U and V be independent $\mathcal{N}(0, 1)$ random variables. Define $Y = V$ and let

$$X \;=\; \begin{cases} U & \text{if } UV \geq 0 \\ -U & \text{if } UV < 0 \end{cases}$$

(a) Let $A \subseteq [0, \infty)$ be a Borel set. Show that $\mathbb{P}(X \in A) = \mathbb{P}(U \in A)$. Hint: Begin with the decomposition $\mathbb{P}(X \in A) = \mathbb{P}(X \in A, UV \geq 0) + \mathbb{P}(X \in A, UV < 0)$.

(b) Carry out a similar analysis for sets $A \subseteq (-\infty, 0)$. Use this and the previous step to show that X has a $\mathcal{N}(0, 1)$ distribution.

(c) Show that $XY = |UV| \geq 0$ and that $\mathrm{Corr}(X, Y) = 2/\pi < 1$. Conclude from these facts that X and Y are not jointly normal.

(d) Show that X^2 and Y^2 are independent.

11. Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be two sequences of random variables defined on the same probability space such that $X_n \sim \mathcal{N}(0,1)$ and $Y_n \sim \mathcal{N}(0,n)$ are independent. Show that $X_n = o_P(Y_n)$.

12. Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be random vectors, possibly defined on different probability spaces, such that $X_n \Rightarrow c$ where $c \in \mathbb{R}^d$ is constant. Show that $X_n \to c$ in probability. Hint: Note that for $\delta > 0$, $I(||x-c|| > \delta) \leq f_\delta(x)$ where $f_\delta(x) = \delta^{-1}||x-c|| \wedge 1$.

13. Let $X \sim \mathcal{N}_n(0, I)$ and $Y \sim \mathcal{N}_n(0, I)$ be independent multinormal random variables. For $0 \leq \theta \leq \pi/2$ define random vectors

$$X(\theta) \;=\; X \sin\theta \,+\, Y \cos\theta$$

$$\dot{X}(\theta) \;=\; X \cos\theta \,-\, Y \sin\theta$$

(a) Show that for each θ, $X(\theta)$ and $\dot{X}(\theta)$ have the same distribution as X.

(b) Show that for each θ, $X(\theta)$ and $\dot{X}(\theta)$ are independent.

14. *Concentration for norms of Gaussian random vectors.* Let $Y \sim \mathcal{N}_d(0, \Sigma)$ and consider the random variable $U = ||Y||$.

(a) Show that $U = F(X)$ in distribution, where $X \sim \mathcal{N}_d(0, I)$ and $F(x) = ||\Sigma^{1/2}x||$

(b) Show that F Lipschitz with constant

$$L \;\leq\; \sup_{u \in \mathbb{R}^d} \frac{||\Sigma^{1/2}u||}{||u||}$$

(c) Find a bound on the right hand side of the inequality above involving the largest eigenvalue of Σ.

(d) Find a concentration inequality for U.

15. Explain and prove the relation $o_p(O_p(1)) = o_p(1)$ for random variables.

16. Let $X_1, \ldots, X_n$ be i.i.d. Exp(1) random variables.

(a) Write down the joint density of $X = (X_1, \ldots, X_n)$ using indicator functions to capture the fact that the variables X_i are positive.

(b) For $1 \le k \le n$ define the random variable $Y_k = X_1 + \cdots + X_k$. Use the general change of variables formula to find the density of $Y = (Y_1, \ldots, Y_n)$.

17. Let $W_n \sim \chi^2_n$ be a chi-squared random variable with n degrees of freedom, and let $\chi^2_{n,\alpha}$ be the upper $1-\alpha$ percentile of the χ^2_n distribution.

(a) Following the arguments in class, find $\mathbb{E}W_n$ and $\text{Var}(W_n)$, and show that

$$\frac{W_n - \mathbb{E}W_n}{\text{Var}(W_n)^{1/2}} \Rightarrow \mathcal{N}(0,1)$$

(b) Use part (a) of the problem to establish the (non-stochastic) relation

$$\frac{\chi^2_{n,\alpha} - n}{\sqrt{n}} \to \sqrt{2} z_\alpha$$

where z_α is the $1-\alpha$ upper percentile of the standard normal. Hint: If the desired result fails to hold, then there is a subsequence $\{n_k\}$ along which the centered and scaled percentiles converge to a number greater than, or less than, $\sqrt{2} z_\alpha$. Use this to get a contradiction.

18. Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be i.i.d. random vectors with $\mathbb{E}X_i = \mu$ and $\text{Var}(X_i) > 0$. Let

$$T_n^2 = (n-1)(\overline{X}_n - \mu)^t S_n^{-1} (\overline{X}_n - \mu)$$

be Hotelling's T^2 statistic, where $S_n = n^{-1} \sum_{i=1}^n (X_i - \overline{X}_n)(X_i - \overline{X}_n)^t$. Show as carefully as you can that $T_n^2 \Rightarrow \chi^2_d$.

19. Let the sample correlation coefficient r_n of a bivariate data set be defined as in class. Show that $1 \le r_n \le 1$.

20. Let X be a random variable with a finite variance and let $Y = \min(X, c)$ for some constant c. Show that the variance of Y exists and is less that or equal the variance of X. [Hint: By considering $Y - c$, show that the assertion is valid for every c if it is valid for $c = 0$. For the case $c = 0$, express X in terms of Y and $Z = \max(X, 0)$, and then consider the covariance of Y and Z.]

21. Let $X_1, X_2, \ldots, X \in \mathbb{R}$ be i.i.d. random variables and let $\mathcal{F}$ be a family of functions $f : \mathbb{R} \to [0,1]$. We say that a *uniform law of large numbers* holds for $\mathcal{F}$ if

$$\sup_{f \in \mathcal{F}} \left| n^{-1} \sum_{i=1}^n f(X_i) - \mathbb{E}f(X) \right| \to 0 \ \text{ wp1 as } n \to \infty \tag{2}$$

a. Show carefully that the Glivenko-Cantelli theorem proved in class is a special case of (2) in which $\mathcal{F} = \{\mathbb{I}_{(-\infty,t]} : t \in \mathbb{R}\}$ is the family of indicator functions of left-infinite closed intervals in $\mathbb{R}$.

b. Use the Glivenko-Cantelli theorem to establish a uniform law of large numbers for the family $\mathcal{F} = \{\mathbb{I}_{(a,b]} : a, b \in \mathbb{R}\}$

c. Show that (2) does not hold when the distribution of X has a density and $\mathcal{F}$ contains the indicator function of every open subset of $\mathbb{R}$.

22. Let $X_1, X_2, \ldots, X \in \mathbb{R}$ be i.i.d. with finite fourth moment, mean μ, variance σ^2, and kurtosis equal to zero. Assume that μ is known. Let

$$\hat{\theta}_n \;=\; \frac{1}{n+2} \sum_{i=1}^{n} (X_i - \mu)^2$$

be an estimator of σ^2 based on $X_1, \ldots, X_n$.

a. If the X_i are normally distributed, what is the minimum mean squared error of an unbiased estimator $\tilde{\theta}_n$ of σ^2 based on $X_1, \ldots, X_n$.

b. Show that $\hat{\theta}_n$ is biased, and find its bias.

c. Find a simple expression for the mean squared error of $\hat{\theta}_n$. Compare this to the lower bound you found in part (a).

23. Let X be a non-negative random variable such that $\mathbb{E}X^2$ is finite. Show that for each $0 < \lambda < 1$ we have the inequality

$$\mathbb{P}(X \geq \lambda\, \mathbb{E}X) \;\geq\; (1-\lambda)^2\, \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}$$

Hint: Use the Cauchy-Schwartz inequality and the identity $X = X\,\mathbb{I}(X \geq c) + X\,\mathbb{I}(X < c)$.

24. 1. Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with $\mathbb{E}X_i = p_i$. Let $S = X_1 + \cdots + X_n$ and let $\mu = \mathbb{E}S = \sum_{i=1}^n p_i$. Use Chernoff's bound and a MGF computation to show that for all $t > \mu$

$$\mathbb{P}(S > t) \leq \exp\{t - \mu - t \log(t/\mu)\}$$

How does this bound compare to Hoeffding's inequality?

25. Let $S(x_1^n : \mathcal{A}) = |\{A \cap \{x_1, \ldots, x_n\} : A \in \mathcal{A}\}|$ be the shatter coefficient of a family $\mathcal{A} \subseteq 2^{\mathcal{X}}$. Show that for every sequence $x_1, \ldots, x_{m+n} \in \mathcal{X}$ we have the sub-multiplicative relation

$$S(x_1^{m+n} : \mathcal{A}) \ \leq \ S(x_1^m : \mathcal{A}) \cdot S(x_{m+1}^{m+n} : \mathcal{A}).$$

26. Let $X \sim \chi_k^2$ have a chi-squared distribution with k degrees of freedom.

(a) Using an identity from a previous homework, or a direct argument, show that if Z is standard normal and $s < 2$ then $\mathbb{E}\exp\{sZ^2\} = (1-2s)^{-1/2}$.

(b) Show that the MGF of X is equal to $\varphi_X(s) = (1-2s)^{-k/2}$.

(c) Use the Chernoff bound and result of Problem 5 above to establish that for $0 \leq \epsilon \leq 1$,

$$P(X \geq (1+\epsilon)k) \leq \exp\left\{-\frac{k}{4}(\epsilon^2 - \epsilon^3)\right\}$$

27. Let X be a random variable and $\mathcal{G}$ a sigma-field such that (i) $\mathbb{E}(X \,|\, \mathcal{G}) = 0$ and (ii) $U \leq X \leq U + c$ with probability one for some $c > 0$ where U is $\mathcal{G}$-measurable. Show that $\mathbb{E}\left[e^{sX} \,|\, \mathcal{G}\right] \leq e^{s^2c^2/8}$ with probability one.

28. Let $H_i(x_1^i) := \mathbb{E}[\, f(X_1^n) \,|\, X_1^i = x_1^i \,]$ be defined as in the proof of McDiarmid's inequality. Show carefully that

$$\sup_{u,u'} [\, H_i(x_1^{i-1}, u) \ - \ H_i(x_1^{i-1}, u') \,] \leq c_i,$$

where c_i is the i'th difference coefficient of f. Note carefully how your argument depends on the independence of $X_1, \ldots, X_n$.

29. *Independent Copies.* Let X, X' be independent random variables with the same distribution. In this case we say that X' is an independent copy of X.

(a) Show that $\text{Var}(X) = \frac{1}{2}\mathbb{E}(X - X')^2$

(b) Argue formally or informally that $\mathbb{E}(X' \,|\, X) = \mathbb{E}X$

(c) Using the result of part (b) and Jensen's inequality for conditional expectations, show that $\mathbb{E}|X - \mathbb{E}X| \leq \mathbb{E}|X - X'|$. This is a key step in establishing a number of important bounds in empirical process theory.

Let $X_1, \ldots, X_n \in \mathcal{X}$ be i.i.d. and let $\mathcal{G}$ be a family of function $g : \mathcal{X} \to [-c, c]$. Define

$$f(x_1^n) \; = \; \sup_{g \in \mathcal{G}} \left| n^{-1} \sum_{i=1}^n g(x_i) \; - \; \mathbb{E} g(X) \right|$$

Find the difference coefficients $c_1, \ldots, c_n$ of f, and use these to establish concentration bounds for the random variable $f(X_1^n)$.

30. Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be independent random vectors such that $\mathbb{E} X_i = 0$ and $||X_i|| \leq c_i/2$ with probability one, where $||u|| = (u^t u)^{1/2}$ is the ordinary Euclidean norm. Let $\alpha = (1/4) \sum_{i=1}^n c_i^2$.

(a) Show that $\mathbb{E} || \sum_{i=1}^n X_i || \leq \sqrt{\alpha}$.

(b) Use the bounded difference inequality and the inequality in part (a) to show that for all $t \geq \sqrt{\alpha}$

$$P \left(|| \sum_{i=1}^n X_i || > t \right) \; \leq \; \exp \left\{ - \frac{(t - \sqrt{\alpha})^2}{2\alpha} \right\}$$

31. Let X be a random variable satisfying the concentration type inequality $\mathbb{P}(|X| > t) \leq a \, e^{-b \, t^2}$ for all $t \geq 0$, where $a \geq 1$ and $b \geq 0$. Show that

$$\mathbb{E}|X| \leq \sqrt{\frac{1 + \log a}{b}}.$$

Hint: Note that for $s \geq 0$ we have $\mathbb{E} X^2 \leq s + \int_s^\infty \mathbb{P}(X^2 \geq t) \, dt$. Use Cauchy-Schwartz.

32. Let $X_1, \ldots, X_n$ be random variables with moment generating functions $\varphi_{X_i}(s) \leq \varphi(s)$ for each $s \geq 0$.

(a) Using the argument in class for Gaussian random variables, show that

$$\mathbb{E} \max(X_1, \ldots, X_n) \; \leq \; \inf_{s > 0} \frac{\log n + \log \varphi(s)}{s}.$$

Suppose now that $U_1, \ldots, U_n$ are Gamma(α, β) random variables.

(b) Show that the moment generating function of U_i is $\varphi(s) = (1 - s \, \beta)^{-\alpha}$.

(c) Using the bound from part (a) and an appropriate choice of s, which can be found by inspection, show that

$$\mathbb{E} \max(U_1, \ldots, U_n) \; \leq \; \frac{2 \, \beta \, \log n}{1 - n^{-1/\alpha}}.$$

33. Let $U_1, \ldots, U_n$ be independent Uniform$(0, \theta)$ random variables. Find $\mathbb{E}\left[\max_{1 \le j \le n} U_j\right]$.

34. Let $V \subseteq \mathbb{R}^n$ be a finite set of vectors $v = (v_1, \ldots, v_n)^t$ with $L = \max_{v \in V} ||v||_2$, and let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher (sign) variables.

(a) Use Hoeffding's MGF inequality to bound the moment generating functions of the random variables $\sum_{i=1}^n \varepsilon_i v_i$ in terms of the constant L.

(b) Show that

$$\mathbb{E}\left[\max_{v \in V} \sum_{i=1}^n \varepsilon_i v_i\right] \;\le\; \sqrt{2L^2 \log |V|}$$

35. Let $\mathcal{X}$ be a set, and let $\mathcal{C} \subseteq 2^{\mathcal{X}}$ be a (possibly infinite) family of sets $C \subseteq \mathcal{X}$. Let $X_1, \ldots, X_n \in \mathcal{X}$ be i.i.d. with distribution μ and define

$$\Delta(X_1^n) \;=\; \sup_{C \in \mathcal{C}} \left| n^{-1} \sum_{i=1}^n I(X_i \in C) \;-\; \mu(C) \right|$$

(a) By carefully adapting the argument for the Symmetrization inequality proved in class, establish directly that

$$\mathbb{E}\Delta(X_1^n) \;\le\; 2\,\mathbb{E} \sup_{C \in \mathcal{C}} \left| n^{-1} \sum_{i=1}^n \varepsilon_i \, I(X_i \in C) \right| \qquad (3)$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent Rademacher (sign) variables. (The idea here is to repeat the arguments in the proof in this special case, *not* to appeal to the general result. Show your work.) The quantity on the right, without the leading factor of two, is sometimes called the expected *Rademacher complexity of* $\mathcal{C}$ with respect to $X_1, \ldots, X_n$.

(b) Show that the Rademacher complexity can be bounded as follows

$$\mathbb{E} \sup_{C \in \mathcal{C}} \left| \sum_{i=1}^n \varepsilon_i \, I(X_i \in C) \right| \;\le\; \sqrt{2n \log \mathbb{E} S(X_1^n : \mathcal{C})}$$

where $S(x_1^n : \mathcal{C}) = |\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\}|$ is the shatter coefficient of the family $\mathcal{C}$. [Hint: First condition on the observations $X_1, \ldots, X_n$.]

(c) Combine the bounds above with the bounded difference inequality to get a high probability bound on $\Delta(X_1^n)$.

36. Show that if $U \sim \chi_n^2$ with $n \ge 3$ then $\mathbb{E}U^{-1} = 1/(n-2)$.

37. Recall that the L_p-norm of a random variable X is defined by $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$. Establish Lyapunov's inequality: If $1 \leq p \leq q$ then $\|X\|_p \leq \|X\|_q$. [Hint: Apply Hölder's inequality with an appropriate choice of conjugate exponents to $|X|^p \cdot 1$.]

38. (Incomplete beta function) Let $\text{Bin}(n,p)$ denote the binomial distribution with parameters $n \geq 1$ and $p \in [0,1]$. Show that for each $1 \leq k \leq n$ and each $p \in [0,1]$ that the following identity holds:

$$P(\text{Bin}(n,p) \geq k) \;=\; \frac{n!}{(k-1)!(n-k)!} \int_0^p u^{k-1}(1-u)^{n-k} du$$

Hint: Fix $1 \leq k \leq n$. Let $f(p)$ and $g(p)$ be, respectively, the left- and right-hand sides of the equation. Show that f, g are equal when $p = 0$. Then show that $f'(p) = g'(p)$ for each $p \in (0,1]$.

39. (Variational characterization of the expected value) Let X be a random variable with finite variance.

(a) Show that $\mathbb{E}X = \arg\min_{a \in \mathbb{R}} \mathbb{E}(X-a)^2$

(b) Show that $\text{Var}(X) = \min_{a \in \mathbb{R}} \mathbb{E}(X-a)^2$

40. (Variational characterization of the median) Let X be a random variable with density f and finite expectation, and let M be a median of X. We wish to establish that

$$M \;=\; \arg\min_{a \in \mathbb{R}} \mathbb{E}|X-a|$$

or equivalently that

$$\mathbb{E}|X-M| \;\leq\; \mathbb{E}|X-a| \text{ for all } a \in \mathbb{R}.$$

a. Replacing X by $X - M$, we may assume without loss of generality that $M = 0$. Let $a > 0$. Express the difference $\mathbb{E}|X-a| - \mathbb{E}|X|$ as a sum of integrals over the disjoint intervals $(-\infty, 0]$, $(0, a]$, and (a, ∞). By carefully considering each integral, show that

$$\mathbb{E}|X-a| - \mathbb{E}|X| \;\geq\; a\,\{\,\mathbb{P}(X \leq 0) - \mathbb{P}(0 < X \leq a) - \mathbb{P}(X > a)\,\}\,.$$

Use the definition of the median and the fact that $a \geq 0$ to conclude that the right side of the inequality above is non-negative. [A similar argument can be carried out for $a \leq 0$, but you do not need to do this.]

b. Suppose now that X has finite variance. Using the variational characterization of the median with $a = \mathbb{E}X$ and Jensen's inequality, show that $|\mathbb{E}X - M| \leq \sqrt{\text{Var}(X)}$.

41. Let X be a random variable taking values in the finite interval $[0, c]$.

(a) Show that $EX \leq c$ and $EX^2 \leq cEX$.

(b) Use these inequalities to show that

$$\text{Var}(X) \;\leq\; c^2[u(1-u)] \quad \text{where} \quad u \;=\; \frac{EX}{c} \in [0,1].$$

(c) Use the result of part (b) to show that $\text{Var}(X) \leq c^2/4$.

(d) Show that this bound is achieved, that is, find a random variable $X \in [0, c]$ for which $\text{Var}(X) = c^2/4$. Hint: put the probability mass of X at the endpoints of the interval.

(e) Use the result in (c) to bound the variance of a random variable X taking values in an interval $[a, b]$ with $-\infty < a < b < \infty$.

42. Let X, Y, and Z be random variables defined on the same probability space, and assume that $\mathbb{E}X^2$ and $\mathbb{E}Y^2$ are finite. Define the conditional covariance of Y and Y given Z by

$$\text{Cov}(X, Y \,|\, Z) \;=\; \mathbb{E}(XY \,|\, Z) - \mathbb{E}(X \,|\, Z)\, \mathbb{E}(Y \,|\, Z).$$

Note that the conditional covariance is a random variable and can be expressed as a function of Z. Use conditioning arguments to establish the following identity, sometimes called the law of total covariance

$$\text{Cov}(X, Y) \;=\; \mathbb{E}\left(\text{Cov}(X, Y \,|\, Z)\right) + \text{Cov}\left(\mathbb{E}(X \,|\, Z), \mathbb{E}(Y \,|\, Z)\right).$$

43. Consider the assertion $o_p(O_p(1)) = o_p(1)$. Provide a rigorous interpretation of the assertion in terms of stochastic sequences, treating the equality as a containment relationship. Establish the assertion as carefully as you can.

44. Let $\mathcal{G}$ be a finite family of functions $g : \mathcal{X} \to [-c, c]$ and let $X_1, \ldots, X_n \in \mathcal{X}$ be iid. Use the union bound and Hoeffding's inequality to find an upper bound on

$$\mathbb{P}\left(\max_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{i=1}^{n} g(X_i) - \mathbb{E}g(X_1) \right| \geq t \right)$$

45. Let $X_1, \ldots, X_n$ be iid $\sim$ Bern(p). Note that $|X_i - p| \leq \max(p, 1-p)$.

(a) Use Bernstein's inequality to get an upper bound on $\mathbb{P}(n^{-1}\sum_{i=1}^n X_i - p \geq t)$ for $t \geq 0$.

(b) Argue that one can restrict attention to $t \in [0, 1-p]$. Using this fact and the bound in part (a) show that if $p \geq 1/2$ then for all $t \geq 0$

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i - p \geq t\right) \leq \exp\left\{\frac{-3nt^2}{8p(1-p)}\right\}$$

(c) Compare the bound in part (b) to a naive inequality based on the central limit theorem and tail bounds for the standard normal distribution.

46. Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be independent random vectors such that $\mathbb{E}X_i = 0$ and $||X_i|| \leq c_i/2$ with probability one, where $||u|| = (u^t u)^{1/2}$ is the ordinary Euclidean norm. Let $\alpha = (1/4)\sum_{i=1}^n c_i^2$.

(a) Show that $\mathbb{E}||\sum_{i=1}^n X_i|| \leq \sqrt{\alpha}$.

(b) Use the bounded difference inequality and the inequality in part (a) to show that for all $t \geq \sqrt{\alpha}$

$$P\left(||\sum_{i=1}^n X_i|| > t\right) \leq \exp\left\{-\frac{(t-\sqrt{\alpha})^2}{2\alpha}\right\}$$

47. Let X and Y be independent random variables with $Y > 0$. Find equalities or inequalities relating the following quantities (you may assume all expecations are finite).

(a) $\mathbb{E}(X/Y)$ and $\mathbb{E}X/\mathbb{E}Y$

(b) $\mathbb{E}Y^3$ and $\mathbb{E}Y\,\mathbb{E}Y^2$

(c) $\mathbb{E}(Y\log Y)$ and $\mathbb{E}Y\log\mathbb{E}Y$

(d) $\mathbb{E}(Y\log Y)$ and $\mathbb{E}Y(\mathbb{E}\log Y)$

48. Let $X_1, \ldots, X_n$ be independent with $\mathbb{E}X = 0$ and $|X_i| \leq c$. Show that if $t \geq n^{-1}\sum_{i=1}^n \mathrm{Var}(X_i)$, then

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \geq t\right) \leq \exp\left\{\frac{-nt}{2+2c/3}\right\}$$

Compare this bound to the one obtained from Hoeffding's inequality in two cases: (i) $\mathrm{Var}(X_i) = 1$ and (ii) $\mathrm{Var}(X_i) = i^{-1}$.

49. Let U_1 and U_2 be independent random variables with mean zero and variance one, and let $U_3 = U_1 + 3U_2$ and $U_4 = 2U_1 - U_2$. Define random vectors $X = (U_1, \ldots, U_4)^t$, $Y = (U_1, U_2)^t$, and $Z = (U_3, U_4)^t$. Note that $X = (Y^t, Z^t)^t$. Find the following

(a) $\text{Var}(X)$

(b) $\text{Var}(Y)$

(c) $\text{Var}(Z)$

(d) $\text{Cov}(Y, Z)$

(e) $\text{Cov}(Z, Y)$

Note that the matrices in (b) - (e) correspond to block submatrices of the variance matrix you found in (a). Which matrices correspond to diagonal blocks, and which matrices correspond to off-diagonal blocks? Discuss.

50. (Bin packing) For $n \geq 1$ let $f_n : [0,1]^n \to \{0, 1, 2, \ldots\}$ be the bin packing function for n objects, that is, $f_n(x_1, \ldots, x_n)$ is the minimum number of length-1 bins needed to hold objects of length $x_1, \ldots, x_n$.

a. Carefully find the difference coefficients $c_1, \ldots, c_n$ of f_n.

b. Let $X_1, \ldots, X_n \in [0,1]$ be independent. Find a bound on $\mathbb{P}(f_n(X_1^n) - \mathbb{E}f_n(X_1^n) \geq t)$ when $t \geq 0$.

c. Now let $x_1, x_2, \ldots \in [0,1]$ and define $a_n = f_n(x_1^n)$. Is the sequence $\{a_n : n \geq 1\}$ subadditive? Justify your answer.

d. What can you say about the limiting behavior of $\mathbb{E}f_n(X_1^n)$ if $X_1, X_2, \ldots \in [0,1]$ is stationary. Justify your answer.

51. Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with $\mathbb{E}X_i = p_i$. Let $S = X_1 + \cdots + X_n$ and let $\mu = \mathbb{E}S = \sum_{i=1}^n p_i$. Use Chernoff's bound and a MGF computation to show that for all $t > \mu$

$$\mathbb{P}(S > t) \leq \exp\{t - \mu - t\log(t/\mu)\}$$

How does this bound compare to Hoeffding's inequality?

52. (Hoeffding's MGF Bound) Let X be a discrete random variable with pmf $p(\cdot)$. Assume that $a \leq X \leq b$ for a, b finite, and that $\mathbb{E}X = 0$. Let $M_X(s) = \mathbb{E}e^{sX}$ be the moment generating function of X and define $\varphi(s) := \log M_X(s)$.

a. Show that
$$\varphi'(s) = \frac{\mathbb{E}[Xe^{sX}]}{\mathbb{E}e^{sX}} \quad \text{and} \quad \varphi''(s) = \frac{\mathbb{E}[X^2e^{sX}]}{\mathbb{E}e^{sX}} - (\varphi'(s))^2$$

b. Verify that $\varphi(0) = \varphi'(0) = 0$

Now fix $t > 0$ and let U be a new random variable having the "exponentially tilted" pmf
$$q(x) \;=\; \frac{p(x)e^{tx}}{\mathbb{E}e^{tX}}$$

c. Verify that $q(\cdot)$ is a pmf and that $a \leq U \leq b$

d. Show that $\mathbb{E}(U) = \varphi'(t)$ and that $\mathrm{Var}(U) = \varphi''(t)$.

e. Using the variance bound for bounded random variables, conclude from (c) and (d) that $\varphi''(t) \leq (b-a)^2/4$.

f. Argue that for $s > 0$, $\varphi(s) \leq s^2(b-a)^2/8$. Exponentiating gives Hoeffding's MGF bound.

53. Carefully reproduce the arguments in class for Bennett's inequality, including the basic MGF bound, and including the details of the Chernoff bound.

54. Let X be a random variable satisfying the concentration type inequality $\mathbb{P}(|X| > t) \leq a\,e^{-b\,t^2}$ for all $t \geq 0$, where $a \geq 1$ and $b \geq 0$. Show that
$$\mathbb{E}|X| \leq \sqrt{\frac{1+\log a}{b}}.$$
Hint: Begin by showing that for $s \geq 0$, $\mathbb{E}X^2 \leq s + \int_s^\infty \mathbb{P}(X^2 \geq t)\,dt$. Use Cauchy-Schwartz.

55. Let $X_1, \ldots, X_n$ be random variables with moment generating functions $M_{X_i}(s) \leq M(s)$ for each $s \geq 0$.

(a) Using the argument in class for Gaussian random variables, show that
$$\mathbb{E}\max(X_1, \ldots, X_n) \;\leq\; \inf_{s>0} \frac{\log n + \log M(s)}{s}.$$

Suppose now that $U_1, \ldots, U_n$ are Gamma(α, β) random variables. Note that the moment generating function of U_i is $M(s) = (1 - s\,\beta)^{-\alpha}$.

(b) Using the bound from part (a) and an appropriate choice of s, which can be found by inspection, show that
$$\mathbb{E}\max(U_1, \ldots, U_n) \;\leq\; \frac{2\,\beta\,\log n}{1 - n^{-1/\alpha}}.$$

56. *Concentration for norms of Gaussian random vectors.* Let $Y \sim \mathcal{N}_d(0, \Sigma)$ and consider the random variable $U = ||Y||$.

(a) Show that $U = F(X)$ in distribution, where $X \sim \mathcal{N}_d(0, I)$ and $F(x) = ||\Sigma^{1/2}x||$

(b) Show that F is Lipschitz with constant

$$L \ \leq \ \sup_{u \in \mathbb{R}^d \setminus \{0\}} \frac{||\Sigma^{1/2}u||}{||u||}$$

(c) Express the right hand side of the inequality above in terms of the eigenvalues of Σ.

(d) Find a concentration inequality for U.

57. Let $\Phi : \mathbb{R} \to (0,1)$ be the CDF of the standard normal, and let $\Phi^{-1} : (0,1) \to \mathbb{R}$ be it's inverse function, equivalently, the percentile function of the standard normal.

(a) Find the limit of $\Phi^{-1}(\alpha)$ as $\alpha \to 0$ and $\alpha \to 1$.

To simplify notation in what follows, let $s(t) = \sqrt{2 \log t}$ for $t \geq 1$.

(b) Use bounds on $\overline{\Phi}(s)$ to show that

$$\lim_{t \to \infty} t \, \overline{\Phi}(s(t)) = 0.$$

(c) Use bounds on $\overline{\Phi}(s)$ to show that for every $\delta \in (0,1)$

$$\lim_{t \to \infty} t \, \overline{\Phi}(\delta \, s(t)) = \infty.$$

(d) Combine the bounds from (b) and (c) to show that

$$\lim_{t \to \infty} \frac{\Phi^{-1}(1 - t^{-1})}{\sqrt{2 \log t}} = 1$$

58. Let $X_1, \ldots, X_n$ be independent standard normal random variables. Here we identify upper and lower bounds for the expectation of $K_n := \max_{1 \leq i \leq n} |X_i|$.

(a) Using the MGF-based bound from class and the fact that $K_n = \max_i(X_i, -X_i)$ show that $\mathbb{E}K_n \leq (2 \log 2n)^{1/2}$.

(b) Let Φ^{-1} be the inverse CDF (percentile function) of the standard normal. Show that

$$K_n \ = \ \Phi^{-1}\left(\frac{1}{2} + \frac{1}{2} \max_{1 \leq i \leq n} V_i\right)$$

where $V_1, \ldots, V_n$ are independent Uniform$(0,1)$ random variables.

(c) Show that $\Phi^{-1}(u)$ is convex on $[1/2, 1)$. Apply Jensen's inequality to the expression in (b) to obtain the bound $\mathbb{E}K_n \geq \Phi^{-1}(1 - 1/(2n+2))$.

(d) Conclude from (a), (c), and the previous problem that $\mathbb{E}K_n/\sqrt{2 \log n} \to 1$ as $n \to \infty$.

59. *Extreme value theory for the Gaussian.* Let a_n and b_n be the extreme value scaling and centering constants for the maximum M_n of n independent standard Gaussian random variables.

(a) Fix $x \in \mathbb{R}$ and let $x_n = x/a_n + b_n$. Show that $n\,\phi(x_n)/x_n \to e^{-x}$ as n tends to infinity. [In your calculations, identify and pay careful attention to the leading order terms.]

(b) Using the result of part (a) and the standard Gaussian tail bound from an earlier homework, show that $n(1 - \Phi(x_n)) \to e^{-x}$.

(c) Use part (b) and the lemma from lecture to show that as n tends to infinity

$$\mathbb{P}(a_n(M_n - b_n) \leq x) \to G(x) = e^{-e^{-x}}$$

(d) Show that $G(x)$ is the CDF of $-\log V$ where $V \sim \text{Exp}(1)$.

60. Let M_n be the maximum of n iid $\mathcal{N}(0,1)$ random variables. Use the Gaussian extreme value theorem to establish the following limiting results.

a. $\mathbb{P}(M_n \geq \sqrt{2 \log n}) \to 0$ as $n \to \infty$

b. $M_n/\sqrt{2 \log n} \to 1$ in probability as $n \to \infty$

61. Let $V \subseteq \mathbb{R}^n$ be a finite set of vectors $v = (v_1, \ldots, v_n)^t$ with $L = \max_{v \in V} ||v||_2$, and let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Rademacher (sign) variables.

(a) Use Hoeffding's MGF inequality to bound the moment generating functions of the random variables $\sum_{i=1}^n \varepsilon_i v_i$ in terms of the constant L.

(b) Show that

$$\mathbb{E}\left[\max_{v \in V} \sum_{i=1}^n \varepsilon_i v_i\right] \;\leq\; \sqrt{2L^2 \log |V|}$$

62. Let X and Y be random variables, possibly defined on different probability spaces, with CDFs F and G, respectively. We say that Y is stochastically larger than X, written $Y \overset{\mathrm{d}}{\geq} X$ if $G(x) \leq F(x)$ for each $x \in \mathbb{R}$. Explain the intuition behind the definition.

(a) Suppose that X, Y are jointly distributed with $X \sim F$, $Y \sim G$, and $Y \geq X$ with probability one. Show that $Y \overset{\mathrm{d}}{\geq} X$.

Let X be a random variable with CDF F. Recall that if F is continuous, then $F(X) \overset{\mathrm{d}}{=} U(0,1)$.

(b) Show that in general, $F(X) \overset{\mathrm{d}}{\geq} U(0,1)$ even if F is not continuous. Hint: Let φ be the percentile function of F. Argue that $F(\varphi(u)) \geq u$ for $0 < u < 1$, and recall that $X \overset{\mathrm{d}}{=} \varphi(U)$ where $U \sim U(0,1)$.

63. *Independent Copies.* Let X, X' be independent random variables with the same distribution. In this case we say that X' is an independent copy of X.

(a) Show that $\mathrm{Var}(X) = \frac{1}{2}\mathbb{E}(X - X')^2$

(b) Argue formally or informally that $\mathbb{E}(X' \,|\, X) = \mathbb{E}X$

(c) Using the result of part (b) and Jensen's inequality for conditional expectations, show that $\mathbb{E}|X - \mathbb{E}X| \leq \mathbb{E}|X - X'|$. This is a key step in establishing a number of important bounds in empirical process theory.

64. Let $X_1, \ldots, X_n$ be iid Rademacher (sign) variables with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$.

(a) Using the variance bound from an earlier HW, show that X_i has maximum variance among all random variables supported on $[-1, 1]$.

(b) Identify the common moment generating function $M_X(s)$ of the X_i, which is a simple sum of exponentials.

(c) Establish the bound $M_X(s) \leq e^{s^2/2}$. Hint: Expand the exponentials in $M_X(s)$, cancel identical terms, and examine the coefficients of the remaining terms.

(d) Use the MGF bound in part (c) and Chernoff's probability bound to find an upper bound on $\mathbb{P}(\sum_{i=1}^n X_i \geq t)$ for $t \geq 0$.

(e) Use Hoeffding's inequality to bound the probability in part (d) and compare the bound you found there. Comment.

65. Let p and q be pmfs on $\{0, 1\}$ with $p(0) = p(1) = 1/2$ and $q(0) = (1 - \epsilon)/2$, $q(1) = (1 + \epsilon)/2$ where $\epsilon \in (0, 1)$. Show that

(a) $\mathrm{KL}(p:q) \;=\; -\frac{1}{2}\log(1-\epsilon^2) \;\leq\; \epsilon^2$ when $\epsilon \;\leq\; \frac{1}{\sqrt{2}}$

(b) $\mathrm{KL}(q:p) \;=\; \frac{1}{2}\log(1-\epsilon^2) + \frac{\epsilon}{2}\log(\frac{1-\epsilon}{1+\epsilon})$

66. (Pinsker's inequality) Pinsker's inequality relates the L_1 distance between two density function to their Kullback-Liebler divergence. It has many uses in statistics and probability. Here we derive Pinsker's inequality from a numerical inequality and Cauchy-Schwartz.

(a) Show that for $x \geq 0$ one has the inequality

$$(x-1)^2 \;\leq\; \left(\frac{4+2x}{3}\right)(x\log x - x + 1)$$

Hint: Let $g(x)$ be the difference between the right- and left-hand sides of the inequality. Expand $g(x)$ in a third order Taylor series around $x = 1$.

(b) Let f and g be probability density functions. Establish Pinsker's inequality

$$\int |f(x) - g(x)|\, dx \;\leq\; \sqrt{2\mathrm{KL}(f:g)}$$

Hint: Note that the left hand side can be written as $\int |f/g - 1|\, g\, dx$. Apply the square root form of the inequality above to the integrand and then apply Cauchy-Schwarz.

67. Let $\mathcal{X}$ be a finite set and let p and q be pmfs on $\mathcal{X}$.

a. Show that $\mathrm{KL}(p:q)$ is infinite if and only if there is some $x \in \mathcal{X}$ with $q(x) = 0$ and $p(x) > 0$. (This simple relation does *not* hold when $\mathcal{X}$ is infinite.)

Let $f : \mathcal{X} \to \mathbb{R}$ be any function on $\mathcal{X}$. Define a new pmf on $\mathcal{X}$ by "exponentially tilting" q according to f as follows

$$q_f(x) \;=\; \frac{e^{f(x)}\, q(x)}{\mathbb{E}_q(e^f)}$$

where $C_f = \sum_{x\in\mathcal{X}} e^{f(x)}\, q(x) > 0$ is the normalizing constant needed to make $q_f(x)$ sum to one. Note that $C_f = \mathbb{E}_q(e^f)$, where $\mathbb{E}_q$ denotes expectation under q.

b. Show that if $\mathrm{KL}(p:q)$ is finite then we have the elementary identity

$$\mathrm{KL}(p:q) - \mathbb{E}_p(f) \;=\; \mathrm{KL}(p:q_f) - \log(C_f)$$

c. Use the previous identity to show that for all p, q we have the following variational expression for the KL divergence:

$$\mathrm{KL}(p:q) \;=\; \sup_{f:\mathcal{X}\to\mathbb{R}} \left[\, \mathbb{E}_p(f) - \log(C_f)\, \right]$$

Hint: Consider separately the case where $\mathrm{KL}(p:q) < \infty$ and $\mathrm{KL}(p:q) = \infty$. In the former case, first establish an inequality, and then find a function f achieving equality.

d. Use the variational expression above to show that the KL divergence is convex, namely, for all pmfs p_1, p_2, q_1, q_2 and all $\alpha \in [0,1]$ we have

$$\mathrm{KL}(\alpha\, p_1 + (1-\alpha)\, p_2 : \alpha\, q_1 + (1-\alpha)\, q_2) \;\leq\; \alpha\, \mathrm{KL}(p_1 : q_1) + (1-\alpha)\, \mathrm{KL}(p_2 : q_2)$$

Hint: Extensive calculations are not necessary.

68. Show that the Kolmogorov-Smirnov distance $\mathrm{KS}(P,Q)$ and the total variation distance $\mathrm{TV}(P,Q)$ are metrics.

69. Establish the log-sum inequality: If $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are positive then

$$\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \;\geq\; \left(\sum_{i=1}^n a_i\right) \log \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}$$

with equality iff all the ratios a_i/b_i are equal. Hint: Use Jensen's inequality and the strict concavity of the log function. Optional: Show that the inequality continues to hold if we assume only that $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are non-negative

70. (Data Processing Inequality) Let p and q be probability mass functions on a countable set $\mathcal{X}$.

(a) Use the log-sum inequality to show that for every event $A \subseteq \mathcal{X}$

$$\sum_{x \in A} p(x) \log \frac{p(x)}{q(x)} \;\geq\; P(A) \log \frac{P(A)}{Q(A)}$$

with equality iff $p(x)/q(x)$ is constant for $x \in A$.

(b) Now let $f : \mathcal{X} \to \mathcal{Y}$ be a function from $\mathcal{X}$ to some other set $\mathcal{Y}$, and let $Y = f(X)$. Find the probability mass function $\tilde{p}$ of Y when $X \sim p$. Find the probability mass function $\tilde{q}$ of Y when $X \sim q$.

(c) Show that $\mathrm{KL}(\tilde{p}, \tilde{q}) \leq \mathrm{KL}(p, q)$

71. (Tensorization) Let $P_1, \ldots, P_n$ and $Q_1, \ldots, Q_n$ be distributions on $\mathbb{R}$ with densities $f_1, \ldots, f_n$ and $g_1, \ldots, g_n$ respectively. Establish the following relations

(a) $\mathrm{KS}(\otimes_{i=1}^n P_i, \otimes_{i=1}^n Q_i) \le \sum_{i=1}^n \mathrm{KS}(P_i, Q_i)$

(b) $\mathrm{TV}(\otimes_{i=1}^n P_i, \otimes_{i=1}^n Q_i) \le \sum_{i=1}^n \mathrm{TV}(P_i, Q_i)$

(c) $\mathrm{KL}(\otimes_{i=1}^n P_i, \otimes_{i=1}^n Q_i) = \sum_{i=1}^n \mathrm{KL}(P_i, Q_i)$

72. Stirling's approximation for factorials states the following

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}$$

(a) Use Stirling's approximation to show that for $s \le n/2$

$$\binom{n}{s} \le \exp\left\{s \log\left(\frac{en}{s}\right)\right\}$$

(b) Let $h(p) = -p \log p - (1-p)\log(1-p)$ for $p \in [0,1]$ be the binary entropy function. Use Stirling's approximation to show that for $s \le n/2$

$$\binom{n}{s} \le 2^{n\,h(s/n)}$$