Statistics 522, Problem Set 3 Solutions

Wellner; 1/230/2004

1. For independent random variables $X_1, \ldots, X_n$, show that

$$P(\max_{1\le i\le n} |X_i| > x) \ge \frac{\sum_1^n P(|X_i| > x)}{1 + \sum_1^n P(|X_i| > x)}, .$$

In particular, if the left side is less than 1/2, then

$$2P(\max_{1\le i\le n} |X_i| > x) \ge \sum_{i=1}^n P(|X_i| > x) .$$

Solution: For $t \ge 0$ we have $1 - t \le \exp(-t)$ and $1 - \exp(-t) \ge t/(1+t)$. Taking $t = \sum_1^n P(|X_i| > x)$ in the first inequality gives

$$\begin{aligned}
\frac{\sum_1^n P(|X_i| > x)}{1 + \sum_1^n P(|X_i| > x)} &\le 1 - \exp\left(-\sum_1^n P(|X_i| > x)\right) \\
&= 1 - \prod_{i=1}^n \exp(-P(|X_i| > x)) \\
&\le 1 - \prod_{i=1}^n (1 - P(|X_i| > x) \\
&\qquad \text{by taking } \ t = P(|X_i| > x) \ \text{ in the first inequality} \\
&= 1 - \prod_{i=1}^n P(|X_i| \le x) \\
&= 1 - P(|X_i| \le x, \ i = 1, \ldots, n) \qquad \text{by independence} \\
&= 1 - P(\max_{1\le i\le n} |X_i| \le x) \\
&= P(\max_{1\le i\le n} |X_i| > x) .
\end{aligned}$$

When the left side of the first inequality is less than 1/2, then we have $1/2 \ge t/(1+t)$ where $t = \sum_1^n P(|X_i| > x)$, and this implies that $t \le 1$, or $1 + t \le 2$ or $1/(1+t) \ge 1/2$. This gives the claimed second inequality. □

2. For $r > 0$, suppose that $X_1, \ldots, X_n$ is a sequence of positive independent random variables with $E|X_i|^r < \infty$ for each i. Let $t_0 \equiv \inf\{t > 0 : \ \sum_1^n P(X_i > t) \le \lambda\}$. Then

$$E \max_{1\le i\le n} |X_i|^r \begin{cases} \le t_0^r + \sum_{i=1}^n \int_{t_0}^\infty P(X_i > t) d(t^r) \\ \ge \frac{\lambda}{1+\lambda} t_0^r + \frac{1}{1+\lambda} \sum_{i=1}^n \int_{t_0}^\infty P(X_i > t) d(t^r) . \end{cases}$$

Hint: Use problem 1.

Solution: First note that

$$E \max_{1\le i\le n} |X_i|^r = E\left(\max_{1\le i\le n} |X_i|\right)^r = \int_0^\infty P(\max_{1\le i\le n} |X_i| > t)d(t^r)\,.$$

To prove the upper bound we split the region of integration into $(0, t_0)$ and $[t_0, \infty)$ and bound the probability by 1 on the first set and by $\sum_1^n P(|X_i| > t)$ on the second set:

$$\begin{aligned} E \max_{1\le i\le n} |X_i|^r &= \left(\int_0^{t_0} + \int_{t_0}^\infty\right) P(\max_{1\le i\le n} |X_i| > t)d(t^r) \\ &\le t_0^r + \int_{t_0}^\infty \sum_{i=1}^n P(|X_i| > t)d(t^r)\,. \end{aligned}$$

To prove the lower bound we use problem #1. Thus

$$\begin{aligned} E \max_{1\le i\le n} |X_i|^r &\ge \int_0^\infty \frac{\sum_1^n P(|X_i| > x)}{1 + \sum_1^n P(|X_i| > x)} d(t^r) \\ &= \left(\int_0^{t_0} + \int_{t_0}^\infty\right) \frac{\sum_1^n P(|X_i| > x)}{1 + \sum_1^n P(|X_i| > x)} d(t^r) \\ &\ge \int_0^{t_0} \frac{\lambda}{1+\lambda} d(t^r) + \int_{t_0}^\infty \frac{1}{1+\lambda} \sum_1^n P(|X_i| > x)d(t^r) \\ &\qquad \text{by the definition of} \quad t_0 \\ &= \frac{\lambda}{1+\lambda} t_0^r + \frac{1}{1+\lambda} \int_{t_0}^\infty \sum_1^n P(|X_i| > x)d(t^r)\,. \end{aligned}$$

3. Let $\epsilon_1, \ldots, \epsilon_n$ be independent Rademacher random variables; i.e. $P(\epsilon_i = \pm 1) = 1/2$. Suppose that $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. (a) Show that

$$P(|\sum_{i=1}^n a_i \epsilon_i| > x) \le 2 \exp\left(-\frac{x^2}{2\|a\|^2}\right)$$

for all $x > 0$; here $\|a\|^2 = \sum_1^n a_i^2$. This is known as *Hoeffding's inequality*. Hint: Use the fact that $(e^y + e^{-y})/2 \le \exp(y^2/2)$ by writing out the series expansions of both sides.
(b) Use the exponential bound in (a) to show that

$$E|\sum_{i=1}^n a_i \epsilon_i|^4 \le 16\|a\|^4\,.$$

This is a somewhat cruder version of the inequality we established in the course of proving the lower bound part of Khinchine's inequality: in that proof we established

$$E|\sum_{i=1}^{n} a_i \epsilon_i|^4 \le 3\|a\|^4$$

by direct calculation.

(c) Use the exponential bound in (a) to show that with $Y \equiv \sum_1^n a_i \epsilon_i$ we have

$$E\exp(tY^2) \le 1 + 2\frac{t}{\left(\frac{1}{2\|a\|^2} - t\right)}$$

for $t < 1/(2\|a\|^2)$.

Solution: (a) For $t > 0$ we have, using $(e^y + e^{-y})/2 \le \exp(y^2/2)$,

$$\begin{aligned}
P(\sum_1^n a_i \epsilon_i > x) &= P(\exp(t\sum_1^n a_i \epsilon_i) > \exp(tx)) \\
&\le e^{-tx} E\exp\left(t\sum_1^n a_i \epsilon_i\right) \\
&= e^{-tx}\prod_{i=1}^{n} E\exp(ta_i\epsilon_i) \\
&= e^{-tx}\prod_{i=1}^{n} \frac{1}{2}(e^{ta_i} + e^{-ta_i}) \\
&\le e^{-tx}\prod_{i=1}^{n} \exp(t^2 a_i^2/2) \\
&= e^{-tx}\exp(t^2\|a\|^2/2) = \exp(-tx + t^2\|a\|^2/2)\,.
\end{aligned}$$

Since this bound holds for every $t > 0$ we can minimize over t: Note that with $c^2 \equiv \|a\|^2$ we have

$$-tx + t^2c^2/2 = \frac{1}{2}(ct - x/c)^2 - \frac{1}{2}\frac{x^2}{c^2} \ge -\frac{1}{2}\frac{x^2}{c^2}$$

with equality if and only if $t = x/c^2$. Thus we find that

$$P(\sum_1^n a_i\epsilon_i > x) \le \exp(-x^2/(2c^2)) = \exp\left(-\frac{x^2}{2\|a\|^2}\right)\,.$$

Repeating this inequality with a replaced by $-a$ and combining we find that

$$P(|\sum_1^n a_i\epsilon_i| > x) \le 2\exp(-x^2/(2c^2)) = \exp\left(-\frac{x^2}{2\|a\|^2}\right)\,.$$

(b) Let $Y \equiv \sum_1^n a_i \epsilon_i$. Then, with $c^2 \equiv \|a\|^2$,

$$\begin{aligned}
E(Y^4) &= 4\int_0^\infty t^3 P(|Y| > y)dy \\
&\le 4\int_0^\infty t^3 2\exp\left(-\frac{y^2}{2c^2}\right)\, dy \\
&= 8\int_0^\infty c^3 v^3 \exp(-v^2/2)\, c dv \\
&\qquad \text{by the change of variables } \ v^2 = y^2/c^2, \text{ or } v = y/c\,, \\
&= 8c^4 \int_0^\infty v^3 \exp(-v^2/2) dv \\
&= 16c^4 \int_0^\infty u \exp(-u) du \\
&\qquad \text{by the change of variables } \ u = v^2/2, \\
&= 16c^4 = 16\|a\|^4\,.
\end{aligned}$$

(c) Much as in our proof of Kolmogorov's exponential lower,

$$\exp(ty^2) = \int_{-\infty}^{y^2} te^{tv}\, dv = \int_{-\infty}^{\infty} te^{tv} 1_{[v \le y^2]} dv\,.$$

Replacing y by $Y \equiv \sum_1^n a_i \epsilon_i$ yields

$$\begin{aligned}
E\exp(tY^2) &= E\left\{\int_{-\infty}^{\infty} te^{tv} 1_{[v \le Y^2]} dv\right\} \\
&= \int_{-\infty}^{\infty} te^{tv} P(|Y| \ge \sqrt{v}) dv \\
&\le 1 + \int_0^\infty te^{tv} 2\exp(-v/(2\|a\|^2)) dv \qquad \text{by the inquality proved in (a)} \\
&= 1 + 2\int_0^\infty t\exp\left(-v\left(\frac{1}{2\|a\|^2} - t\right)\right)\, dv \\
&= 1 + 2t\frac{1}{\frac{1}{2\|a\|^2} - 1} \qquad\qquad \text{if } \ t < 1/(2\|a\|^2)\,.
\end{aligned}$$

4. Exercise 11.7.3, PfS, page 289.

Solution: (a) Let $\epsilon > 0$. Now by Markov's inequality

$$\limsup_{n\to\infty} P(|X_n| > M) \le \frac{\limsup_{n\to\infty} E|X_n|^r}{M^r} < \epsilon$$

for $M > M(r,\epsilon) \equiv (\limsup_{n\to\infty} E|X_n|^r/\epsilon)^{1/r}$. Thus there is an $N \equiv N(\epsilon, r)$ such that $\sup_{n>N} P(|X_n| > M) \le 2\epsilon$. But since $F_1, \ldots, F_N$ are df's, there exists a $K = K_\epsilon$ so large that $\max_{1\leq n\leq N} P(|X_n| > K) < 2\epsilon$. Taking $R = R_\epsilon = \max\{M, K\}$ we have

$$\sup_{1\leq n<\infty} P(|X_n| > R) < 2\epsilon\,.$$

Thus $\{F_n\}$, the family of distributions of $\{X_n\}$, is tight.
(b) Let $\epsilon > 0$. Let $r = r(\epsilon) \in C_F$ be so large that $1 - F(r) < \epsilon/4$, and let $l = l(\epsilon) \in C_F$ be so small that $F(l) < \epsilon/4$. Now there exists an $N = N_\epsilon$ so large that

$$|F_n(r) - F(r)| < \epsilon/4 \qquad \text{for all} \qquad n > N$$

and

$$|F_n(l) - F(l)| < \epsilon/4 \qquad \text{for all} \qquad n > N\,.$$

Then we have

$$\begin{aligned}\sup_{n>N} F_n([l,r]^c) &\leq F(l) + (1 - F(r)) \\ &\qquad + |F_n(l) - F(l)| + |F_n(r) - F(r)| \\ &\leq \epsilon/4 + \epsilon/4 + \epsilon/4 + \epsilon/4 = \epsilon\end{aligned}$$

But since $F_1, \ldots, F_N$ are distribution functions, we can easily find an interval $[l', r']$, such that

$$\max_{1\leq n\leq N} F_n([l', r']^c) < \epsilon\,,$$

and hence the interval $[a, b] \equiv [l \wedge l', r \vee r']$ satisfies

$$\sup_{1\leq n<\infty} F_n([a,b]^c) < \epsilon\,;$$

i.e $\{F_n\}$ is tight.