Sta-230: Probability **Spring 2017**

Supplemental notes for topic 9: April 4, 6

9.1 Polynomial inequalities

Theorem (Jensen)**.** *If ϕ is a convex function then $\phi(\mathbb{E}x) \leq \mathbb{E}\phi(x)$.*

Theorem (Bienaymé-Chebyshev)**.** *For any random variable x, $\epsilon > 0$*

$$\mathbf{P}(|x| \geq \epsilon) \leq \frac{\mathbb{E}x^2}{\epsilon^2}.$$

Proof

$$\mathbb{E}x^2 \geq E(x^2 I_{\{|x| \geq \epsilon\}}) \geq \epsilon^2 \mathbf{P}(|x| > \epsilon). \ \square$$

Theorem (Markov)**.** *For any random variable x, $\epsilon > 0$*

$$\mathbf{P}(|x| \geq \epsilon) \leq \frac{\mathbb{E}e^{\lambda x}}{e^{\lambda \epsilon}}$$

and

$$\mathbf{P}(|x| \geq \epsilon) \leq \inf_{\lambda < 0} e^{-\lambda \epsilon} \mathbb{E}e^{\lambda x}.$$

Proof

$$\mathbf{P}(x > \epsilon) = \mathbf{P}(e^{\lambda x} > e^{\lambda \epsilon}) \leq \frac{\mathbb{E}e^{\lambda x}}{e^{\lambda \epsilon}}. \ \square$$

9.2 Exponential inequalities

For the sums or averages of independent random variables the above bounds can be improved from polynomial in $1/\epsilon$ to exponential in ϵ.

Theorem (Bennet)**.** *Let $x_1, ..., x_n$ be independent random variables with $\mathbb{E}x = 0$, $\mathbb{E}x^2 = \sigma^2$, and $|x_i| \leq M$. For $\epsilon > 0$*

$$\mathbf{P}\left(|\sum_{i=1}^n x_i| > \epsilon\right) \leq 2e^{\frac{-n\sigma^2}{M^2}\phi\left(\frac{\epsilon M}{n\sigma^2}\right)},$$

where

$$\phi(z) = (1+z)\log(1+z) - z.$$

Proof

We will prove a bound on one-side of the above theorem

$$\mathbf{P}\left(\sum_{i=1}^n x_i > \epsilon\right).$$

$$\mathbf{P}\left(\sum_{i=1}^{n} x_i > \epsilon\right) \le e^{-\lambda\epsilon}\mathbb{E}e^{\lambda\sum x_i} = e^{-\lambda\epsilon}\Pi_{i=1}^{n}\mathbb{E}e^{\lambda x_i}$$
$$= e^{-\lambda\epsilon}(\mathbb{E}e^{\lambda x})^n.$$

$$\mathbb{E}e^{\lambda x} = \mathbb{E}\sum_{k=0}^{\infty}\frac{(\lambda x)^k}{k!} = \sum_{k=0}^{\infty}\lambda^k\frac{\mathbb{E}x^k}{k!}$$
$$= 1 + \sum_{k=2}^{\infty}\frac{\lambda^k}{k!}\mathbb{E}x^2x^{k-2} \le 1 + \sum_{k=2}^{\infty}\frac{\lambda^k}{k!}M^{k-2}\sigma^2$$
$$= 1 + \frac{\sigma^2}{M^2}\sum_{k=2}^{\infty}\frac{\lambda^k M^k}{k!} = 1 + \frac{\sigma^2}{M^2}(e^{\lambda M} - 1 - \lambda M)$$
$$\le e^{\frac{\sigma^2}{M^2}(e^{\lambda M} - \lambda M - 1)}.$$

The last line holds since $1 + x \le e^x$.

Therefore,

$$\mathbf{P}\left(\sum_{i=1}^{n} x_i > \epsilon\right) \le e^{-\lambda\epsilon}e^{\frac{\sigma^2}{M^2}(e^{\lambda M} - \lambda M - 1)}. \tag{9.1}$$

We now optimize with respect to λ by taking the derivative with respect to λ

$$\begin{aligned} 0 &= -\epsilon + \frac{n\sigma^2}{M^2}(Me^{\lambda M} - M), \\ e^{\lambda M} &= \frac{\epsilon M}{n\sigma^2} + 1, \\ \lambda &= \frac{1}{M}\log\left(1 + \frac{\epsilon M}{n\sigma^2}\right). \end{aligned}$$

The theorem is proven by substituting λ into equation (9.1). $\square$

The problem with Bennet's inequality is that it is hard to get a simple expression for ϵ as a function of the probability of the sum exceeding ϵ.

Theorem (Bernstein)**.** *Let $x_1, ..., x_n$ be independent random variables with $\mathbb{E}x = 0$, $\mathbb{E}x^2 = \sigma^2$, and $|x_i| \le M$. For $\epsilon > 0$*

$$\mathbf{P}\left(|\sum_{i=1}^{n} x_i| > \epsilon\right) \le 2e^{-\frac{\epsilon^2}{2n\sigma^2 + \frac{2}{3}\epsilon M}}.$$

Proof

Take the proof of Bennet's inequality and notice

$$\phi(z) \ge \frac{z^2}{2 + \frac{2}{3}z}. \ \square$$

Remark. *With Bernstein's inequality a simple expression for ϵ as a function of the probability of the sum exceeding ϵ can be computed*

$$\sum_{i=1}^{n} x_i \leq \frac{2}{3}uM + \sqrt{2n\sigma^2 u}.$$

$$\mathbf{P}\left(\sum_{i=1}^{n} x_i > \epsilon\right) \leq 2e^{-\frac{\epsilon^2}{2n\sigma^2+\frac{2}{3}\epsilon M}} = e^{-u},$$

where

$$u = \frac{\epsilon^2}{2n\sigma^2 + \frac{2}{3}\epsilon M}.$$

we now solve for ϵ

$$\epsilon^2 - \frac{2}{3}\epsilon M - 2n\sigma^2\epsilon = 0$$

and

$$\epsilon = \frac{1}{3}uM + \sqrt{\frac{u^2M^2}{9} + 2n\sigma^2 u}.$$

Since $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$

$$\epsilon = \frac{2}{3}uM + \sqrt{2n\sigma^2 u}.$$

So with large probability

$$\sum_{i=1}^{n} x_i \leq \frac{2}{3}uM + \sqrt{2n\sigma^2 u}. \ \triangle$$

If we want to bound

$$|n^{-1}\sum_{i=1}^{n} f(x_i) - \mathbb{E}f(x)|$$

we consider

$$|f(x_i) - \mathbb{E}f(x)| \leq 2M.$$

Therefore

$$\sum_{i=1}^{n}(f(x_i) - \mathbb{E}f(x)) \leq \frac{4}{3}uM + \sqrt{2n\sigma^2 u}$$

and

$$n^{-1}\sum_{i=1}^{n} f(x_i) - \mathbb{E}f(x) \leq \frac{4}{3}\frac{uM}{n} + \sqrt{\frac{2\sigma^2 u}{n}}.$$

Similarly,

$$\mathbb{E}f(x) - n^{-1}\sum_{i=1}^{n} f(x_i) \geq \frac{4}{3}\frac{uM}{n} + \sqrt{\frac{2\sigma^2 u}{n}}.$$

In the above bound

$$\sqrt{\frac{2\sigma^2 u}{n}} \geq \frac{4uM}{n}$$

which implies $u \leq \frac{n\sigma^2}{8M^2}$ and therefore

$$|n^{-1}\sum_{i=1}^{n} f(x_i) - \mathbb{E}f(x)| \lesssim \sqrt{\frac{2\sigma^2 u}{n}} \text{ for } u \lesssim n\sigma^2,$$

which corresponds to the tail probability for a Gaussian random variable and is predicted by the Central Limit Theorem (CLT) Condition that $\lim_{n\to\infty} n\sigma^2 \to \infty$. If $\lim_{n\to\infty} n\sigma^2 = C$, where C is a fixed constant, then

$$|n^{-1}\sum_{i=1}^{n} f(x_i) - \mathbb{E}f(x)| \lesssim \frac{C}{n}$$

which corresponds to the tail probability for a Poisson random variable.

We now look at an even simpler exponential inequality where we do not need information on the variance.

Theorem (Hoeffding)**.** *Let $x_1, ..., x_n$ be independent random variables with $\mathbb{E}x = 0$ and $|x_i| \le M_i$. For $\epsilon > 0$*

$$\mathbf{P}\left(|\sum_{i=1}^{n} x_i| > \epsilon\right) \le 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^{n} M_i^2}}.$$

Proof

$$\mathbf{P}\left(\sum_{i=1}^{n} x_i > \epsilon\right) \le e^{-\lambda\epsilon}\mathbb{E}e^{\lambda\sum_{i=1}^{n} x_i} = e^{-\lambda\epsilon}\Pi_{i=1}^{n}\mathbb{E}e^{\lambda x_i}.$$

It can be shown

$$\mathbb{E}(e^{\lambda x_i}) \le e^{\frac{\lambda^2 M_i^2}{8}}.$$

The bound is proven by optimizing the following with respect to λ

$$e^{-\lambda\epsilon}\Pi_{i=1}^{n} e^{\frac{\lambda^2 M_i^2}{8}}. \ \square$$

Applying Hoeffding's inequality to

$$n^{-1}\sum_{i=1}^{n} f(x_i) - \mathbb{E}f(x)$$

we can state that with probability $1 - e^{-u}$

$$n^{-1}\sum_{i=1}^{n} f(x_i) - \mathbb{E}f(x) \le \sqrt{\frac{2Mu}{n}},$$

which is a sub-Gaussian as in the CLT but without the variance information we can never achieve the $\frac{1}{n}$ rate we achieved when the random variable has a Poisson tail distribution.

We will use the following version of Hoeffding's inequality in later lectures on Kolmogorov chaining and the Dudley's entropy integral.

Theorem (Hoeffding)**.** *Let $x_1, ..., x_n$ be independent random variables with $\mathbf{P}(x_i = M_i) = 1/2$ and $\mathbf{P}(x_i = -M_i) = 1/2$. For $\epsilon > 0$*

$$\mathbf{P}\left(|\sum_{i=1}^{n} x_i| > \epsilon\right) \le 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^{n} M_i^2}}.$$

Proof

$$\mathbf{P}\left(\sum_{i=1}^{n} x_i > \epsilon\right) \leq e^{-\lambda\epsilon}\mathbb{E}e^{\lambda\sum_{i=1}^{n} x_i} = e^{-\lambda\epsilon}\Pi_{i=1}^{n}\mathbb{E}e^{\lambda x_i}.$$

$$\begin{aligned}
\mathbb{E}(e^{\lambda x_i}) &= \frac{1}{2}e^{\lambda M_i} + \frac{1}{2}e^{-\lambda M_i}, \\
\frac{1}{2}e^{\lambda M_i} + \frac{1}{2}e^{-\lambda M_i} &= \sum_{k=0}^{\infty}\frac{(M_i\lambda)^{2k}}{(2k)!} \leq e^{\frac{\lambda^2 M_i^2}{2}}.
\end{aligned}$$

Optimize the following with respect to λ

$$e^{-\lambda\epsilon}\Pi_{i=1}^{n}e^{\frac{\lambda^2 M_i^2}{2}}. \quad \square$$

9.3 Strong law of large numbers

Lemma 9.1 (Borel-Cantelli)**.** *Consider a set of events $\{A_n : n \geq 1\}$ in a probability space. If*

$$\sum_{n=1}^{\infty}\mathbf{P}(A_n) < \infty$$

then $\mathbf{P}(A_n \textit{occurs infinitely often}) = 0$.

The idea behind the Borel-Cantelli Lemma is that if an event happens finitely often and the state space is infinite then the probability of the event heppening is zero. So we can say something happens with probability zero.

We now use the Borel-Cantelli Lemma and the weak law of large numbers to prove the strong law of large numbers.

Consider an infinite sequence of Bernoulli random variable $X_1, ..., X_n \overset{iid}{\sim}$ Be(p). We want to show that $\mathbf{P}(\lim_{n\to\infty}\frac{S_n}{n} = p) = 1$.

If we can show that $\mathbf{P}(|S_n| > n\varepsilon \text{ i.o.}) = 0$ then $\mathbf{P}(\lim_{n\to\infty}\frac{S_n}{n} = p) = 1$ because these are complementary events.

To make notation easier we consider the case where $p = 1/2$ the generalization is straightforward. Consider the event

$$A_n = \{\omega \in \Omega : |S_n| \geq n\varepsilon\},$$

We have an exponential inequality for the above event since we know by Hoeffding's inequality

$$\begin{aligned}
\mathbf{P}\left(|\sum_{i=1}^{n} x_i| > n\varepsilon\right) &\leq 2e^{-\frac{2n^2\varepsilon^2}{\sum_{i=1}^{n} 1^2}}, \\
\mathbf{P}\left(|\sum_{i=1}^{n} x_i| > n\varepsilon\right) &\leq 2e^{-2n\varepsilon^2}.
\end{aligned}$$

So we now check if for all $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) \leq \sum_{n=1}^{\infty} 2e^{-2n\varepsilon^2} < \infty.$$

The sum of exponential tails is bounded so Borel-Cantelli holds and we see that for Bernoulli random variables the strong law of large numbers holds.