EE5110: Probability Foundations for Electrical Engineers **July-November 2015**

Lecture 27: Concentration Inequalities

Lecturer: Dr. Krishna Jagannathan *Scribe: Arjun Nadh*

A concentration inequality is a result that gives us a probability bound on certain random variables taking atypically large or atypically small values. While concentration of probability measures is a vast topic, we will only discuss some foundational concentration inequalities in this lecture.

27.1 Markov's Inequality

If X is a non-negative random variable, with $\mathbb{E}[X] < \infty$, then for any $\alpha > 0$,

$$\mathbb{P}(X > \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}.$$

Clearly, this inequality is meaningful only when $\alpha > \mathbb{E}[X]$.

Proof:

$$\begin{aligned}\mathbb{E}[X] &\overset{(a)}{=} \mathbb{E}[X\mathbb{I}_{\{X\leq\alpha\}}] + \mathbb{E}[X\mathbb{I}_{\{X>\alpha\}}],\\ &\overset{(b)}{\geq} \mathbb{E}[X\mathbb{I}_{\{X>\alpha\}}],\\ &\geq \alpha\mathbb{P}(X > \alpha).\end{aligned}$$

where (a) follows from linearity of expectations. Since X is a non-negative random variable $\mathbb{E}[X\mathbb{I}_{\{X\leq\alpha\}}] \geq 0$ and thus (b) follows. ■

Markov Inequality is probably the most fundamental concentration inequality, although it is usually quite loose. After all, the bound decays rather slowly, as $1/\alpha$. Tighter bounds can be derived under stronger assumptions on the random variable. For example, when the variance is finite, we have Chebyshev's inequality.

27.2 Chebyshev Inequality

If X is a random variable with expectation μ and variance $\sigma^2 < \infty$, then

$$\mathbb{P}(| X - \mu | > k\sigma) \leq \frac{1}{k^2}, \quad k > 0.$$

This can also be written as

$$\mathbb{P}(| X - \mu | > c) \leq \frac{\sigma^2}{c^2}, \quad c > 0.$$

Proof: The proof follows by applying Markov's inequality to the non-negative random variable $\mid X - \mu \mid^2$.

$$\begin{aligned} \mathbb{P}(\mid X-\mu\mid^2 > (k\sigma)^2) &\leq \frac{\mathbb{E}(\mid X-\mu\mid^2)}{(k\sigma)^2}, \\ &= \frac{\sigma^2}{(k\sigma)^2}, \\ &= \frac{1}{k^2}, \\ \Rightarrow \mathbb{P}(\mid X-\mu\mid > (k\sigma)) &\leq \frac{1}{k^2}. \end{aligned}$$

■

Note that the Chebyshev's bound decays as $1/k^2$, an improvement over the basic Markov inequality. As one might imagine, exponentially decaying bounds can be derived by invoking the Markov inequality, as long as the moment generating function exists in a neighbourhood of the origin. This result is known as the Chernoff bound, which we present briefly.

27.3 Chernoff Bound

Let $M_X(s) = \mathbb{E}[e^{sX}]$ and assume that $M_X(s) < \infty$ for $s \in [-\epsilon, \epsilon]$ for some $\epsilon > 0$. Then

$$\mathbb{P}(X > \alpha) \leq e^{-\Lambda^*(\alpha)},$$

where $\Lambda^*(\alpha) = \sup_{s>0}(s\alpha - \log M_X(s))$.

Proof: For any $s > 0$,

$$\mathbb{P}(X > \alpha) = \mathbb{P}(e^{sX} > e^{s\alpha}).$$

By Markov's Inequality,

$$\begin{aligned} \mathbb{P}(X > \alpha) &\leq \frac{\mathbb{E}[e^{sX}]}{e^{s\alpha}}, \\ \mathbb{P}(X > \alpha) &\leq M_X(s)e^{-s\alpha}, \quad \forall s > 0 \text{ and } s \in D_X, \end{aligned} \tag{27.1}$$

where $D_X = \{s \mid M_X(s) < \infty\}$.

In (27.1), note that the bound decays exponentially in α for *every* $s > 0$ belonging to D_X. The tightest such exponential bound is obtained by infimising the right hand side:

$$\begin{aligned} \mathbb{P}(X > \alpha) &\leq \inf_{s>0} M_X(s)e^{-s\alpha}, \\ &= e^{-\sup\limits_{s>0}(s\alpha - \log M_X(s))}. \end{aligned}$$

Thus

$$\mathbb{P}(X > \alpha) \leq e^{-\Lambda^*(\alpha)}.$$

■

This gives us an exponentially decaying bound for the 'positive tail' $\mathbb{P}(X > \alpha)$. Similarly we can prove a Chernoff bound for the negative tail $\mathbb{P}(X < \alpha)$ by taking $s < 0$.

27.4 Exercise

1. Let X_1, X_2, ..., X_n be i.i.d. random variables with PDF f_X. Then the set of random variables X_1, X_2, ..., X_n is called a *random sample* of size n of X. The sample mean is defined as

$$\overline{X_n} = \frac{1}{n}(X_1 + X_2 + ... + X_n).$$

   Let X_1, X_2,.. X_n be a random sample of X with mean μ and variance σ^2. How many samples of X are required for the probability that the sample mean will not deviate from the true mean μ by more than $\sigma/10$ to be at least .95?

2. A biased coin, which lands heads with probability $\frac{1}{10}$ each time it is flipped, is flipped 200 times consecutively. Give an upper bound on the probability that it lands heads at least 120 times.

3. A post-office handles 10,000 letters per day with a variance of 2,000 letters. What can be said about the probability that this post office handles between 8,000 and 12,000 letters tomorrow? What about the probability that more than 15,000 letters come in?