1 The Law of Large Numbers

A general rule in probability is that the aggregate behavior of random events become more predictable when we have many independent sources of randomness. Suppose that we run an experiment where we flip a coin n times. If we record the number of heads then we expect that

$$\frac{\#\text{ of heads}}{n} \approx \frac{1}{2}.$$

The more coins we flip, the closer the proportion of heads will be to 0.5. Of course, the outcomes are random so it might never be exactly 0.5, but one expects that somehow the events become less random the more experiments we run. This intuition can be made precise, and this is called the law of large numbers.

1.1 Law of Large Numbers

Consider i.i.d. random variables $X_1, \dots, X_n$ with mean μ and variance σ^2. Recall that the *sample mean* is given by

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i.$$

Using the properties of the expected and variance of linear combinations, it follows that

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu \quad \text{and} \quad \mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$

This suggests that the variance of the sample mean shrinks as $n \to \infty$. The weak law of large number says that the probability the sample mean is not equal to its expected value goes to zero.

Theorem 1 (*Weak Law of Large Numbers*)

Let $X_1, \dots, X_n$ be i.i.d. *with mean μ and variance σ^2. Then for any $\epsilon > 0$,*

$$\mathbb{P}(|\bar{X}_n - \mu| > \epsilon) \to 0 \text{ as } n \to \infty.$$

In other words, $\bar{X}_n \to \mu$ in probability.

In fact, we have a stronger statement that doesn't even need to assume that the variance is finite. The probability that the sample mean takes the value μ in the limit is exactly 1.

Theorem 2 (*Strong Law of Large Numbers*)

Let $X_1, \dots, X_n$ be i.i.d. *with mean μ. Then*

$$\mathbb{P}(\{\bar{X}_n(\omega) \to \mu\}) = 1$$

In other words, $\bar{X}_n \to \mu$ almost surely.

Remark 1. There are many notions of the convergence of random variables. We have seen convergence in distribution already, and we have just introduced the notions of convergence in probability and almost sure convergence. We have that almost sure convergence implies convergence in probability, so it is a stronger notion of convergence. The reverse implication is not true in general.

1.2 Connection Between Tail Probabilities and Moments

The moments of the random variable can be used to control the probability that a random variable takes extreme outlying values away from its expected value.

Theorem 3 (*Markov's Inequality*)

For any random variable and constant $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{t}.$$

Remark 2. This is intuitive, because a small expected value means that the probability that $|X|$ is big should be small. Furthermore, the more extreme the outlier is (which is measured by t), the more rare it should be. For example, if X denotes income of a randomly selected person in a city, then

$$\mathbb{P}(X \geq 2\,\mathbb{E}[X]) \leq \frac{1}{2}$$

since we can't have more than half the people in the city making twice the average income.

This is a good first step to controlling the tail events since it only requires some information about the first moment $\mathbb{E}[|X|]$. However, if we have more information such as control of the second moment, then we have better control of the tails.

Corollary 1 (*Chebyshev's Inequality*)

Let X *have mean* μ. *Then for any constant* $a > 0$,

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\mathbb{E}[(X-\mu)^2]}{a^2} = \frac{\mathrm{Var}(X)}{a^2}.$$

Remark 3. Notice that the tail probability in Chebyshev's Inequality is now bounded by t^{-2} instead of t^{-1} so it gives us better control than Markov's inequality.

If we have even more information about the moments, then we can get upgrade the bound to give us exponential control of the tails.

Corollary 2 (*Chernoff Bound*)

For any random variable and constantss $t > 0$ *and* $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}} = M_X(t)e^{-ta}.$$

There is an extra parameter t in the statement of the Chernoff bound, so we can minimize the upper bound as a function of t to get the better control of the tail probabilities.

1.3 Example Problems

Problem 1.1. An insurance company sells $1,000$ independent health insurance policies. Suppose that for any policy holder, the expected claim is $5,000$ but the variance is $2,000,000$ since the costs of the treatment can vary a lot.

1. What can the company say about the average claim over all $1,000$ policies?
2. What is a bound on the probability that the insurance company has to pays more than $10,000,000$ in a given year.

Solution 1.1.

Part 1: The law of large number states that the average claim

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X_1] = 5000.$$

So, since $n = 1000$ is quite a large number of policies, the average claim will be close to 5000 even though the individual claim amounts can vary widely.

Part 2: Notice that $T = \sum_{i=1}^{1000} X_i$ the total total claim amount satisfies

$$\mathbb{E}(T) = \sum_{i=1}^{1000} \mathbb{E}[X_i] = 5,000,0000 \text{ and } \operatorname{Var}(T) = \sum_{i=1}^{1000} \operatorname{Var}\Big(X_i\Big) = 2,000,000,000.$$

Therefore, by Chebyshev's inequality

$$\begin{aligned}\mathbb{P}(T > 10,000,0000) &= \mathbb{P}(T - 5,000,000 > 5,000,000)\\ &\leq \mathbb{P}(|T - 5,000,000| > 5,000,000)\\ &\leq \frac{\operatorname{Var}(T)}{5,000,000^2} = \frac{2,000,000,000}{5,000,000^2} = 0.0008.\end{aligned}$$

Remark 4. If we used Markov's inequality instead, we would have gotten

$$\mathbb{P}(T > 10,000,0000) \leq \frac{\mathbb{E}(T)}{10,000,000} = \frac{1}{2}$$

which is true, but not as good of a bound as if we used information about the variance.

Problem 1.2. Suppose a machine produces metal rods whose lengths (in cm) are nonnegative random variables L with mean 100.

1. Use Markov's inequality to bound the probability that a rod exceeds 150 cm.
2. Suppose that we now know that the variance of the rods are 25 cm^2. Use Chebyshev's inequality to bound the probability that a rod exceeds 150 cm.

Solution 1.2.

Part 1: By Markov's inequality (since the lengths of the rods are non-negative)

$$\mathbb{P}(L > 150) \leq \frac{\mathbb{E}[L]}{150} = \frac{2}{3}.$$

Part 2: By Markov's inequality (since the lengths of the rods are non-negative)

$$\mathbb{P}(L > 150) \leq \frac{\mathbb{E}[L]}{150} = \frac{2}{3}.$$

Problem 1.3. Let Y be the average income (in thousands) of 100 independently surveyed people from a population with mean 50 and variance 100. Use Chebyshev's inequality to bound

$$\mathbb{P}(|Y - 50| \geq 10).$$

Solution 1.3. We have that Y is the average of, so

$$\operatorname{Var}(Y) = \operatorname{Var}\left(\frac{1}{100}\sum_{i=1}^{100} X_i\right) = \frac{100\operatorname{Var}(X)}{100^2} = 1$$

Therefore,

$$\mathbb{P}(|Y - 50| \geq 10) \leq \frac{\operatorname{Var}(Y)}{10^2} = \frac{1}{100},$$

which is quite small even though the standard deviation of a single person is 10.

1.4 Proofs of Key Results

Problem 1.4. Prove Theorem 1, the weak law of large number.

Solution 1.4. We have that $\operatorname{Var}(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}(X_i) = \frac{\sigma^2}{n}$. For any $\epsilon > 0$, Chebyshev's inequality implies that

$$\mathbb{P}(|\bar{X}_n - \mu| > \epsilon) = \frac{\operatorname{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2 n} \to 0$$

as $n \to \infty$.

Problem 1.5. Prove Theorem 3, Markov's inequality.

Solution 1.5. For any $a > 0$, we define $Y = \frac{|X|}{a}$. It suffices to show that $\mathbb{P}(Y > 1) \leq \mathbb{E}[Y]$ since

$$\mathbb{P}(|X| > a) = \mathbb{P}(Y > 1) \leq \mathbb{E}[Y] = \frac{\mathbb{E}[|X|]}{a}.$$

Notice that we always have

$$\mathbb{1}(Y \geq 1) = \begin{cases} 1 & Y \geq 1 \\ 0 & Y < 1 \end{cases} \leq Y.$$

Monotonicity of the expected value implies that

$$\mathbb{E}[\mathbb{1}(Y \leq 1)] \leq \mathbb{E}[Y] \implies \mathbb{P}(Y > 1) \leq \mathbb{E}[Y].$$

Problem 1.6. Prove Corollary 1 and Corollary 2. That is, for any $t > 0$ and $a > 0$, show that

1. $$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\operatorname{Var}(X)}{a^2}$$
2. $$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$$

Solution 1.6. Both of these statements are direct consequences of Markov's inequality.

Part 1: We have

$$\mathbb{P}(|X - \mu| \geq a) = \mathbb{P}(|X - \mu|^2 \geq a^2),$$

so Markov's inequality implies that

$$\mathbb{P}(|X-\mu|^2 \geq a^2) \leq \frac{\mathbb{E}[|X-\mu|^2]}{a^2} = \frac{\mathrm{Var}(X)}{a^2}.$$

Part 2: We have

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}),$$

so Markov's inequality implies that

$$pP(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}} = M_X(t)e^{-ta}.$$

2 The Central Limit Theorem