Independent random variables

E6711: Lectures 3
Prof. Predrag Jelenković

1 Last two lectures

- probability spaces
- probability measure
- random variables and stochastic processes
- distribution functions
- independence
- conditional probability
- memoriless property of geometric and exponential distributions
- expectation
- conditional expectation (double expectation)
- mean-square estimation

Let $\{X_j, j \geq 1\}$ be a sequence of independent random variables and

$$N_n = \sum_{j=1}^{n} X_j$$

be a partial sum of the first n of these r.v.s. In many applications understanding the statistical behavior of these sums is very important. Thus, a big part of probability theory studies the characteristics of N_n.

In this lecture we review some of the well-known theorems of probability theory:

- Markov and Chebyshev's inequalities
- Laws of Large Numbers
- Central Limit Theorem

2 Inequalities

Proposition 2.1 (Markov's inequality) *If X is a nonnegative random variable, then for any $a > 0$*

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}X}{a}.$$

Proof: For $a > 0$, let us define an indicator function

$$1[X \geq a] = \begin{cases} 1 & \text{if} \quad X \geq a \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$1[X \geq a] \leq \frac{X}{a};$$

thus, by taking the expected value on both sides in the preceding inequality we obtain

$$\mathbb{E}1[X \geq a] = \mathbb{P}[X \geq a] \leq \frac{\mathbb{E}X}{a}.$$

$\diamond$

Corollary 2.1 (Chebyshev's inequality) *If X is a random variable with finite mean μ and variance $\sigma^2 = \mathbb{E}(X-\mu)^2$, then for any $\epsilon > 0$*

$$\mathbb{P}[|X-\mu| \geq \epsilon] \leq \frac{\sigma^2}{\epsilon^2}.$$

Proof: Let $Y = (X-\mu)^2$, then

$$\begin{aligned}
\mathbb{P}[|X-\mu| \geq \epsilon] &= \mathbb{P}[(X-\mu)^2 \geq \epsilon^2] \\
&= \mathbb{P}[Y \geq \epsilon^2] \\
&\leq \frac{\mathbb{E}Y}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2},
\end{aligned}$$

where the last inequality follows from Markov's inequality. $\diamond$

Corollary 2.2 (Chernoff's bound) *Let $M(t) \stackrel{\text{def}}{=} \mathbb{E}e^{tX} < \infty$ for some $t > 0$, then*

$$\mathbb{P}[X \geq y] \leq e^{-ty}M(t).$$

Proof: Let $Y = e^{tX}$ and $a = e^{ty}$, then

$$\begin{aligned}
\mathbb{P}[X \geq y] &= \mathbb{P}[tX \geq ty] \\
&= \mathbb{P}[e^{tX} \geq e^{ty}] \\
&= \mathbb{P}[Y \geq a] \\
&\leq \frac{\mathbb{E}Y}{a} = e^{-ty}M(t);
\end{aligned}$$

note that the last inequality follows from Markov's inequality. $\diamond$

3 Laws of Large Numbers: ergodic theorems

Ergodic theory studies the conditions under which the sample path average

$$Y \stackrel{\text{def}}{=} \frac{X_1 + \cdots + X_n}{n} \tag{3.1}$$

converges to the mean $\mu = \mathbb{E}X_1$ as $n \to \infty$.

Theorem 3.1 (Weak Law of Large Numbers) *Let $X_1, X_2, \ldots$, be a sequence of independent random variables with finite mean μ and variance σ^2. Then, for any $\epsilon > 0$*

$$\mathbb{P}\left[\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right] \to 0 \quad \text{as} \quad n \to \infty.$$

Proof: Recall the definition of Y from equation (3.1), then

$$\mathbb{E}Y = \mathbb{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = \mu$$

and

$$\begin{aligned}
\text{Var}(Y) &= \text{Var}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\
&= \frac{\text{Var}(\text{X}_1) + \cdots + \text{Var}(X_n)}{n^2} \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

Thus, by Chebyshev's inequality

$$\mathbb{P}[|Y - \mu| \geq \epsilon] \leq \frac{\sigma^2}{n\epsilon^2} \to 0 \quad \text{as} \quad n \to \infty,$$

we conclude the proof of the theorem. $\diamond$

Now we know that

$$\mathbb{P}\left[\left|\frac{X_1 + X_2 + \cdots + X_n}{n}\right| \geq \epsilon\right]$$

converges to zero, however it is not clear how fast? This problem is investigated by the theory of **Large Deviations**.

Recall $M(t) = \mathbb{E}e^{tX_1}$ and define the *rate function*

$$l(a) \stackrel{\text{def}}{=} -\log\left(\inf_{t\geq 0} e^{-ta} M(t)\right) = \sup_{t\geq 0}(ta - \log M(t)).$$

Then

Theorem 3.2 *For every $a > \mathbb{E}X_1$ and $n \geq 1$*

$$\mathbb{P}\left[\frac{X_1 + X_2 + \cdots + X_n}{n} \geq a\right] \leq e^{-nl(a)}.$$

Proof: For any $t > 0$

$$\begin{aligned}
\mathbb{P}\left[\frac{X_1 + X_2 + \cdots + X_n}{n} \geq a\right] &= \mathbb{P}\left[e^{t(X_1+X_2+\cdots+X_n)} \geq e^{tan}\right] \\
\text{(Chernoff's inequality)} \quad &\leq e^{-tan}\mathbb{E}e^{t(X_1+X_2+\cdots+X_n)} \\
&= \left(e^{-ta}\mathbb{E}e^{tX_1}\right)^n.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{P}\left[\frac{X_1 + X_2 + \cdots + X_n}{n} \geq a\right] &\leq \inf_{t\geq 0}\left(e^{-ta}\mathbb{E}e^{tX_1}\right)^n \\
&= e^{-nl(a)}.
\end{aligned}$$

$\diamond$

The preceding two theorems estimate the probabilities that a sample path mean is close to the (ensemble) mean. The following theorem goes one step further in showing that for almost every *fixed* omega the sample path average converges to the mean (in the ordinary deterministic sense).

Theorem 3.3 (Strong Law of Large Numbers) *Let* $X_1, X_2, \ldots$*, be a sequence of independent random variables with finite mean* μ *and* $K \stackrel{\text{def}}{=} \mathbb{E}X_1^4 < \infty$*. Then, for almost every* ω *(or with probability 1)*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \quad \text{as} \quad n \to \infty.$$

Remark: For this theorem to hold it is enough to assume that the mean $\mu = \mathbb{E}X_1$ exists (i.e., it could be even infinite). However, in order to present a simpler proof, we impose a stronger assumption $\mathbb{E}X_1^4 < \infty$.

Proof: To begin, assume that $\mu = \mathbb{E}X_j = 0$; then

$$\mathbb{E}N_n^4 = \mathbb{E}[(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)(X_1 + \cdots + X_n)].$$

Now, expanding the right-hand side of the equation above will result in terms of the form ($i \neq j \neq k$)

$$\begin{aligned}
&\mathbb{E}X_i^4 \\
&\mathbb{E}[X_i^3 X_j] = \mathbb{E}X_i^3 \mathbb{E}X_j = 0 \quad \text{by independence} \\
&\mathbb{E}X_i^2 X_j^2 \\
&\mathbb{E}[X_i^2 X_j X_k] = \mathbb{E}X_i^2 \mathbb{E}X_j \mathbb{E}X_k = 0 \quad \text{by independence} \\
&\mathbb{E}[X_i X_j X_k X_l] = \mathbb{E}X_i \mathbb{E}X_j \mathbb{E}X_k \mathbb{E}X_l = 0 \quad \text{by independence.}
\end{aligned}$$

Next, there are n terms of the form $\mathbb{E}X_i^4$ and for each $i \neq j$ there are $\binom{4}{2} = 6$ terms in the expansion that are equal to $\mathbb{E}X_i^2 X_j^2$. Hence,

$$\begin{aligned} \mathbb{E}N_n^4 &= n\mathbb{E}X_1^4 + 6\binom{n}{2}(\mathbb{E}X_1^2)^2 \\ &= nK + 3n(n-1)(\mathbb{E}X_1^2)^2. \end{aligned} \tag{3.2}$$

Also, $K < \infty$ implies $\mathbb{E}X_1^2 < \infty$, since

$$0 \leq \mathrm{Var}(X_1^2) = \mathbb{E}X_1^4 - (\mathbb{E}X_1^2)^2 \quad \Rightarrow \quad (\mathbb{E}X_1^2)^2 \leq K \tag{3.3}$$

Now, by replacing (3.3) in (3.2), we obtain

$$\frac{\mathbb{E}N_n^4}{n^4} \leq \frac{K}{n^3} + \frac{3K}{n^2} \leq \frac{4K}{n^2}.$$

Thus,

$$\mathbb{E}\sum_{n=1}^{\infty} \frac{N_n^4}{n^4} = \sum_{n=1}^{\infty} \frac{\mathbb{E}N_n^4}{n^4} \leq \sum_{n=1}^{\infty} \frac{4K}{n^2} < \infty.$$

Therefore, with probability 1

$$\sum_{n=1}^{\infty} \frac{N_n^4}{n^4} < \infty,$$

which implies that, with probability 1

$$\lim_{n\to\infty} \frac{N_n^4}{n^4} = 0,$$

or equivalently

$$\mathbb{P}\left[\lim_{n\to\infty} \frac{N_n}{n} = 0\right] = 1.$$

This concludes the proof of the case $\mu = 0$. If $\mu \neq 0$, then define $X_j' = X_j - \mathbb{E}X_j$ and use the same proof. $\diamond$

4 Central Limit Theorem

Central Limit Theorem, Similarly to the Large Deviation Theorem, measures the deviation of the sample mean from the expected value μ.

Theorem 4.1 (Cental Limit Theorem (CLT)) *Let $X_j, j \geq 1$ be a sequence of i.i.d. r.v.s with mean μ and variance $\sigma^2 < \infty$. Then, the distribution of*

$$Z_n \stackrel{\text{def}}{=} \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to standard normal distribution as $n \to \infty$, i.e., for any real number a

$$\mathbb{P}\left[\frac{X_1 + \cdots X_n - n\mu}{\sigma\sqrt{n}} \leq a\right] \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx \quad \text{as} \quad n \to \infty.$$

First, we state the following key lemma that will be used in the proof of CLT.

Lemma 4.1 *Let $Z_1, Z_2, \ldots$, be a sequence of r.v.s having distribution functions F_{Z_n} and moment generating functions $M_{Z_n}(t) = \mathbb{E}e^{tZ_n}, n \geq 1$; And let Z be a random variable having distribution F_Z and moment generating function $M_Z(t)$. It $M_{Z_n}(t) \to M_Z(t)$ as $n \to \infty$, for all t, then*

$$F_{Z_n}(x) \to F_Z(x) \quad \text{as} \quad x \to \infty.$$

Proof: Omitted. $\diamond$

Proof of CLT: Assume that $\mu = 0$ and $\sigma^2 = 1$. Then, moment generating function (m.g.f.) of $X_j/\sqrt{n}$ is equal to

$$\mathbb{E}\left[e^{tX_j/\sqrt{n}}\right] = M(t/\sqrt{n}) \quad \text{where} \quad M(t) = \mathbb{E}e^{tX_j}.$$

Thus, the m.g.f. of $\sum_{j=1}^n X_j/\sqrt{n}$ is equal to

$$\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Now, if $L(t) \stackrel{\text{def}}{=} \log M(t)$, then

$$\log\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = nL\left(\frac{t}{\sqrt{n}}\right) = \frac{L(t/\sqrt{n})}{n^{-1}}.$$

Thus

$$\begin{aligned}
\lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} \quad \text{(by L'Hospital's rule)} \\
&= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})t}{-2n^{-1/2}} \\
&= \lim_{n\to\infty} \frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}} \quad \text{(by L'Hospital's rule)} \\
&= \lim_{n\to\infty} L''(t/\sqrt{n})\frac{t^2}{2}
\end{aligned}$$

Next, note that

$$L(0) = 0 \quad L'(0) = \frac{M'(0)}{M(0)} = \mu = 0$$

$$L''(0) = \frac{M(0)M''(0) - (M'(0))^2}{(M(0))^2} = \mathbb{E}X^2 = 1.$$

Hence, for any finite t

$$\lim_{n\to\infty} L''(t/\sqrt{n}) = 1$$

and, therefore

$$\lim_{n\to\infty} nL(t/\sqrt{n}) = \frac{t^2}{2},$$

or, equivalently

$$\lim_{n\to\infty} \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = e^{t^2/2}.$$

On the other hand, if Z is a standard normal r.v., then

$$\mathbb{E}e^{tN} = e^{t^2/2},$$

which, by Lemma 4.1, concludes the proof of the theorem for $\mu = 0$ and $\sigma^2 = 1$.

For the general case $\mu \neq 0$ and $\sigma^2 \neq 1$, we can introduce new variables

$$X_j^* \stackrel{\text{def}}{=} \frac{X_j - \mu}{\sigma};$$

clearly

$$\mathbb{E}X_j^* = 0 \quad \text{and} \quad \mathrm{Var}(X_j^*) = 1,$$

and, therefore, we can use the already proved case. $\diamond$