Probability 2 - Notes 12

The Law of Large Numbers (LLN)

The LLN is one of the most important results of the classical probability theory. We shall discuss here the so called weak form of this Law.

Theorem 1. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables each with finite mean $E(X_j) = a$ and finite variance $Var(X_j) = \sigma^2$. Then for any $\varepsilon > 0$*

$$P(|\frac{1}{n}\sum_{j=1}^{n} X_j - a| \geq \varepsilon) \to 0 \text{ as } n \to \infty.$$

The proof of this theorem will be given later. In order to carry it out we need the following lemmas.

Lemma 1 (Markov's Inequality). *If ξ is a random variable s. t. $\xi \geq 0$ and $E(\xi) < \infty$ then for any $\delta > 0$ the following inequality holds:*

$$P(\xi \geq \delta) \leq \frac{E(\xi)}{\delta} \tag{1}$$

Proof. Define a r.v. η by setting

$$\eta = \begin{cases} 1 & \text{if } \xi \geq \delta \\ 0 & \text{if } \xi < \delta \end{cases}$$

Note that $\xi \geq \delta\eta$. To see this, consider two cases: $\xi \geq \delta$ and $\xi < \delta$. In the first case the inequality holds because $\xi \geq \delta \equiv \delta\eta$. In the second case the inequality holds because $\xi \geq 0 \equiv \delta\eta$.
But then also $E(\xi) \geq E(\delta\eta) = \delta E(\eta)$. Note that $E(\eta) = P(\eta = 1) = P(\xi \geq \delta)$ and hence $E(\xi) \geq \delta P(\xi \geq \delta)$ which is equivalent to the statement of the Lemma. □

Exercise Prove that For any $h > 0$, $P(|X| \geq h) \leq \frac{E[|X|]}{h}$. Hint: apply (1) to $\xi = |X|$.

Lemma 2: Chebyshev's Inequality. *If $E(\xi) = \mu$ and $Var(\xi) = \sigma^2$, which are finite, then for any $\varepsilon > 0$*

$$P(|\xi - \mu| \geq \varepsilon) \leq \frac{Var(\xi)}{\varepsilon^2}. \tag{2}$$

Proof. Obviously, $P(|\xi - \mu| \geq \varepsilon) = P((\xi - \mu)^2 \geq \varepsilon^2)$. Since $(\xi - \mu)^2 \geq 0$, we can apply (1) with ξ replaced by $(\xi - \mu)^2$ and $\delta = \varepsilon^2$. Hence $P(|\xi - \mu| \geq \varepsilon) \leq \frac{E(\xi-\mu)^2}{\varepsilon^2} = \frac{Var(\xi)}{\varepsilon^2}$. □

Lemma 3. *If $X_1, X_2, \ldots, X_n$ is a sequence of r.v.'s with $Cov(X_j, X_i) = 0$ for all $j \neq i$, then*

$$Var(\sum_{j=1}^{n} X_j) = \sum_{j=1}^{n} Var(X_j). \tag{3}$$

Proof. By the definition of variance, $Var(\sum_{j=1}^{n} X_j) = E[\sum_{j=1}^{n} X_j - E(\sum_{j=1}^{n} X_j)]^2 =$
$E[\sum_{j=1}^{n}(X_j - E(X_j))]^2 = \sum_{j=1}^{n}(X_j - E(X_j))^2 + 2\sum_{1\leq j<i\leq n} E[(X_j - E(X_j))(X_i - E(X_i))] =$
$\sum_{j=1}^{n} Var(X_j) + 2\sum_{1\leq j<i\leq n} Cov(X_j, X_i) = \sum_{j=1}^{n} Var(X_j)$. □

Proof of the LLN (Theorem 1). Set $\xi = \frac{1}{n}\sum_{j=1}^{n} X_j$ and note that $E[\xi] = \frac{1}{n}\sum_{j=1}^{n} E(X_j) = \frac{1}{n}na = a$. Hence, by Chebyshev's inequality (2),

$$P(|\frac{1}{n}\sum_{j=1}^{n} X_j - a| \geq \varepsilon) = P(|\xi - a| \geq \varepsilon) \leq \frac{Var(\frac{1}{n}\sum_{j=1}^{n} X_j)}{\varepsilon^2}$$

Using the properties of the variance and (3) we obtain

$$P(|\frac{1}{n}\sum_{j=1}^{n} X_j - a| \geq \varepsilon) \leq \frac{Var(\sum_{j=1}^{n} X_j)}{n^2\varepsilon^2} = \frac{\sum_{j=1}^{n} Var(X_j)}{n^2\varepsilon^2} = \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0 \text{ as } n \to \infty. \ \square$$

Note. The following definition is useful: we say that a sequence S_n of r.v.'s *converges in probability* to a if $\lim_{n\to\infty} P(|S_n - a| \geq \varepsilon) = 0$ for any $\varepsilon > 0$. Theorem 1 thus states that the sequence $\frac{1}{n}\sum_{j=1}^{n} X_j$ *converges in probability* to a as n tends to infinity.

Bernoulli's Law of Large Numbers.

Theorem 2. *Consider a series of n independent Bernoulli trials and let p be the probability of 'success' in each trial. Denote by ν_n the total number of 'success' in n trials. Then $\frac{\nu_n}{n}$ converges in probability to p as n tends to infinity:*

$$\lim_{n\to\infty} P(|\frac{\nu_n}{n} - p| \geq \varepsilon) = 0 \text{ for any } \varepsilon > 0.$$

Proof. Let ξ_j be the number of successes in the j^{th} trial. Obviously ξ_j are independent r.v.'s such that $P(\xi_j = 1) = p$ and $P(\xi_j = 0) = 1 - p \equiv q$ and $\nu_n = \sum_{j=1}^{n} \xi_j$. Obviously $E(\xi_j) = p$, $Var(\xi_j) = pq$ and, by Theorem 1, $\frac{\nu_n}{n} \equiv \frac{1}{n}\sum_{j=1}^{n} \xi_j$ converges in probability to p as $n \to \infty$. $\square$

In fact the proof of Theorem 1 is based on an estimate which is of its own importance , namely for any $\varepsilon > 0$

$$P(|\frac{\nu_n}{n} - p| \geq \varepsilon) \leq \frac{pq}{n\varepsilon^2}.$$

Some examples using the inequalities.

1. If X is a non-negative random variable with $E(X) = \mu > 0$, then it follows from Markov's inequality with $\delta = N\mu$ that $P(X > N\mu) \leq \frac{\mu}{N\mu} = \frac{1}{N}$ for any $N > 0$.

2. If $\sigma^2 = 0$ then from Chebyshev's inequality for any $h > 0$, $P(|X - \mu| < h) = 1 - P(|X - \mu| \geq h) \geq 1 - \frac{\sigma^2}{h^2} = 1$. Hence $P(X = \mu) = \lim_{h\downarrow 0} P(|X - \mu| < h) = 1$. So variance zero implies the random variable takes a single value with probability 1.

3. When $\sigma^2 > 0$ Chebyshev's inequality gives a lower bound on the probability that X lies within k standard deviations from the mean. Take $\varepsilon = k\sigma$. Then

$$P(|X - \mu| < k\sigma) = 1 - P(|X - \mu| \geq k\sigma) \geq 1 - \frac{\sigma^2}{(k\sigma)^2} = 1 - \frac{1}{k^2}$$

4. When $\sigma = 1$, how large a sample is needed if we want to be at least 95% certain that the sample mean lies within 0.5 of the true mean? We shall use Chebyshev's inequality to estimate

this number. Remember that the sample mean is defined as $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$. By (2) with $\varepsilon = 0.5$ we have

$$P(|\overline{X}_n - a| < 0.5) = 1 - (|\overline{X}_n - \mu| \geq 0.5) \geq 1 - \frac{\sigma^2}{n(0.5)^2} = 1 - \frac{4}{n} \geq 0.95$$

provided $n \geq \frac{4}{0.05} = 80$. So a sample of size 80 would be sufficient for the purpose.

A remark on application to statistics.

Consider a sequence of i.i.d. random variables $X_1, X_2, ...$ and let $Y_n = g(X_1, ..., X_n)$ be an estimator of a parameter θ of the common distribution of the $X's$. If Y_n is an unbiased estimator of θ then $E[Y_n] = \theta$. Let $\sigma_n^2 = Var(Y_n)$. Then for any $\varepsilon > 0$, using Chebyshev's inequality for Y_n, we obtain $P(|Y_n - \theta| \geq \varepsilon) \leq \frac{\sigma_n^2}{\varepsilon}$. This suggests that when comparing unbiased estimators we should choose the one with smallest variance. We would also like our estimator to be as accurate as we please provided we take a large enough sample. If $\lim_{n\to\infty} \sigma_n^2 = 0$ we can ensure this, since $\lim_{n\to\infty} P(|Y_n - \theta| \geq \varepsilon) = 0$ for any $\varepsilon > 0$, i.e. Y_n converges in probability to θ (and Y_n is said to be a *consistent* estimator of θ).

The Central Limit Theorem (CLT).

Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables each with finite mean μ and finite variance σ^2 and let $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ be the sample mean based on $X_1, ..., X_n$. Then we can find an approximation for $P(\overline{X}_n \leq A)$ when n is large by writing the event for X_n in terms of the standardized variable $Z_n = \sqrt{n}(X_n - \mu)/\sigma$ (i.e. $P(\overline{X}_n \leq A) = P\left(Z_n \leq \frac{\sqrt{n}(A-\mu)}{\sigma}\right)$) and proving that $\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ which is the c.d.f. of $N(0,1)$. The proof of this result uses the m.g.f. and the following lemma.

Lemma. *Let $Z_1, Z_2, ...$ be a sequence of random variables. If $\lim_{n\to\infty} M_{Z_n}(t) = M(t)$, which is the m.g.f. of a distribution with c.d.f. F, then $\lim_{n\to\infty} F_{Z_n}(z) = F(z)$ at all points z for which $F(z)$ is continuous.*

Theorem (The Central Limit Theorem). *Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables with finite mean $E(X_j) = \mu$ and finite variance $Var(X_j) = \sigma^2$. Moreover, suppose that X_j have a m.g.f. which exists in an open region about zero. Let*

$$Z_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma \equiv \frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n}\sigma}$$

then

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z).$$

Proof. Let $U_j = (X_j - \mu)/\sigma$ and let $M_U(t)$ be the common m.g.f. Then $M_U(t) = e^{-\mu t/\sigma} M_X(t/\sigma)$ exists in an open interval about $t = 0$, $M(0) = 1$, $M'(0) = E[U] = 0$ and $M''(0) = E[U^2] = Var(U) = 1$. So $U_1, U_2, ...$ are i.i.d. with mean zero and variance one. Now

$$M_{Z_n}(t) = E\left[e^{t\sum_{j=1}^{n} U_j/\sqrt{n}}\right] = \prod_{j=1}^{n} E[e^{tU_j/\sqrt{n}}] = \left(M_U(t/\sqrt{n})\right)^n$$

Taking logs to base e gives $\ln(M_{Z_n}(t)) = n(\ln(M_U(t/\sqrt{n})))$. Now let $x = 1/\sqrt{n}$ and use L'Hopital's rule. Then

$$\begin{aligned}
\lim_{n\to\infty} n\ln(M_U(t/\sqrt{n})) &= \lim_{x\downarrow 0} \frac{\ln(M_U(xt))}{x^2} \\
&= \lim_{x\downarrow 0} \frac{tM_U^{'}(xt)/M_U(xt)}{2x} = \lim_{x\downarrow 0} \frac{t^2(M_U^{''}(xt)M_U(xt) - (M_U^{'}(xt))^2)/(M_U(xt))^2}{2} \\
&= \frac{t^2(M_U^{''}(0)M_U(0) - (M_U^{'}(0))^2)}{2(M_U(0))^2} = \frac{t^2}{2}
\end{aligned}$$

Hence $\lim_{t\to\infty} \ln(M_{Z_n}(t)) = t^2/2$ and so $\lim_{t\to\infty} M_{Z_n}(t) = e^{t^2/2}$. Since this is the m.g.f. of the $N(0,1)$ distribution, using the lemma proves that $\lim_{n\to\infty} P(Z_n \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$.
$\square$