MATHEMATICS 191, FALL 2003-2004
MATHEMATICAL PROBABILITY
Laws of Large Numbers

These proofs are based on material scattered all over the text.

1 Weak Law - simple special case

Random variable X has $\mathbb{P}(X=0)=\mathbb{P}(X=2)=\frac{1}{2}$ so its generating function is $G_X(s)=\frac{1}{2}(1+s^2)$. Then X/n has generating function $G(s)=\frac{1}{2}(1+s^{\frac{2}{n}})$, since its possible values are 0 and $\frac{2}{n}$.

We are interested in the average of n independent copies of X when n is large. This is $\frac{1}{n}S_n$ where $S_n=\sum_{i=1}^n X_n$. For this average the generating function is

$$G_n(s)=G(s)^n=[\frac{1}{2}(1+s^{\frac{2}{n}})]^n$$

.

Now $s^{\frac{2}{n}}=e^{\frac{2\log s}{n}}$. When n is large, the exponent is small (for any positive s), so we can keep just two terms of the Taylor expansion. Then

$$G_n(s)=G(s)^n=[\frac{1}{2}(1+1+\frac{2\log s}{n})+o(\frac{1}{n})]^n.$$

$$G_n(s)=[(1+\frac{\log s}{n})+o(\frac{1}{n})]^n.$$

and

$$\lim_{n\to\infty}G_n(s)=e^{\log s}=s.$$

This is the generating function for a random variable with the constant value 1.

2 Central limit theorem - simple special case

We adjust X so that its mean is 0. Random variable X now has $\mathbb{P}(X=-1)=\mathbb{P}(X=1)=\frac{1}{2}$ so its generating function is $G_X(s)=\frac{1}{2}(s^{-1}+s)$.

Conveniently, the variance of X is 1.

Again we sum n independent copies of X: $S_n=\sum_{i=1}^n X_n$. But now we divide the sum by $\sqrt{n}$ instead of by n. For $\frac{1}{\sqrt{n}}S_n$ the generating function is

$$G_n(s)=[\frac{1}{2}(s^{-\frac{1}{\sqrt{n}}}+s^{\frac{1}{\sqrt{n}}})]^n.$$

or

$$G_n(s)=[\frac{1}{2}(e^{-\frac{\log s}{\sqrt{n}}}+e^{\frac{\log s}{\sqrt{n}}})]^n.$$

Again expand the exponentials in a Taylor series, but keep three terms.

$$e^{\frac{\log s}{\sqrt{n}}} = 1 + \frac{\log s}{\sqrt{n}} + \frac{1}{2}\frac{\log^2 s}{n} + o(\frac{1}{n}).$$

$$G_n(s) = [\frac{1}{2}(1 - \frac{\log s}{\sqrt{n}} + \frac{1}{2}\frac{\log^2 s}{n} + 1 + \frac{\log s}{\sqrt{n}} + \frac{1}{2}\frac{\log^2 s}{n} + o(\frac{1}{n})]^n.$$

The terms in $\log s$ cancel one another and

$$G_n(s) = [1 + \frac{\log^2 s}{2n} + o(\frac{1}{n})]^n.$$

So

$$\lim_{n\to\infty} G_n(s) = e^{\frac{\log^2 s}{2}}.$$

What distribution has this generating function? Try the unit normal distribution $N(0,1)$. Of course, now the random variable can take any real value so we must sum rather than integrating.

$$G_{normal}(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s^x e^{-\frac{x^2}{2}} dx.$$

Set $s^x = e^{x \log s}$ and complete the square in the integrand to obtain

$$G_{normal}(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\log s)^2}{2}} e^{\frac{(\log s)^2}{2}} dx.$$

After the change of variable $u = x - \log s$

$$G_{normal}(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} e^{\frac{\log^2 s}{2}} du = e^{\frac{\log^2 s}{2}}.$$

So $\lim_{n\to\infty} \frac{S_n}{\sqrt{n}}$ has the same generating function as the normal distribution with mean 0 and variance 1. Using the book's approach, with "characteristic functions," we could go a step further and invoke the Fourier inversion theorem to say that the probability distribution functions are also the same.

3 Weak Law - general case

This is a self-contained simplification of the proof on p. 193 of the text.

Assume merely that random variable X has expectation μ. This implies that its generating function $G_X(s)$ is differentiable at $s = 1$ and that $G'_X(1) = \mu$. It follows, by the definition of the derivative, that for s close to 1

$$G_X(s) = G_X(1) + G'_X(1)(s-1) + o(s-1)$$

There is no need to assume that any higher derivatives of $G_X(s)$ exist.

Now $G_X(1) = 1$ and $G'_X(1) = \mu$ so

$$G_X(s) = 1 + \mu(s-1) + o(s-1)$$

For the random variable $\frac{X}{n}$ all the values assumed are smaller by a factor of n and so the generating function is

$$1 + \mu(s^{\frac{1}{n}} - 1) + o(s^{\frac{1}{n}} - 1)$$

Finally, for the sum of n independent copies of this random variable the generating function is

$$G_n(s) = [1 + \mu(s^{\frac{1}{n}} - 1) + o(s-1)]^n = [1 + \mu(e^{\frac{\log s}{n}} - 1) + o(e^{\frac{\log s}{n}} - 1)]^n$$

Keep two terms of the Taylor series for the exponential, and so approximate $e^{\frac{\log s}{n}} - 1$ by $\frac{\log s}{n}$:

$$G_n(s) = [1 + \frac{\mu \log s}{n} + o(\frac{\log s}{n})]^n$$

Take the limit as n approaches infinity:

$$\lim_{n\to\infty} G_n(s) = e^{\mu \log s} = s^\mu$$

which is the generating function for a random variable with the constant value μ.

4 Central limit theorem - general case

This is a self-contained simplification of the proof on page 194 of the text. Assume that the random variable X has an expectation of 0 and a variance of σ^2, which means that $\mathbb{E}(X^2) = \sigma^2$.

Expand its generating function in a Taylor series about $s = 1$:

$$G_X(s) = G_X(1) + G'_X(1)(s-1) + \frac{1}{2}G''_X(1)(s-1)^2 + o((s-1)^2)$$

Identifying $G_X(1) = 1$, $G'_X(1) = 0$ (the mean), $G''_X(1) = \sigma^2$

$$G_X(s) = 1 + \frac{1}{2}\sigma^2(s-1)^2 + o((s-1)^2)$$

Now divide X by $\frac{1}{\sqrt{n\sigma^2}}$ and sum up n independent copies. This gives $\frac{1}{\sqrt{n\sigma^2}}S_n$, where $S_n = \sum_{i=1}^{n} X_n$.

The generating function for $\frac{X}{\sqrt{n\sigma^2}}$ is

$$1 + \frac{1}{2}\sigma^2(s^{\frac{1}{\sqrt{n\sigma^2}}} - 1)^2 + o((s^{\frac{1}{\sqrt{n\sigma^2}}} - 1)^2)$$

The generating function for $\frac{1}{n\sigma^2}S_n$ is

$$G_n(s) = [1 + \frac{1}{2}\sigma^2(s^{\frac{1}{\sqrt{n\sigma^2}}} - 1)^2 + o((s^{\frac{1}{\sqrt{n\sigma^2}}} - 1)^2)]^n$$

But

$$s^{\frac{1}{\sqrt{n\sigma^2}}} = e^{\frac{\log s}{\sqrt{n\sigma^2}}}$$

For large n, $e^{\frac{\log s}{\sqrt{n\sigma^2}}} - 1$ is well approximated by $\frac{\log s}{\sqrt{n\sigma^2}}$ and so

$$G_n(s) = [1 + \frac{1}{2}\sigma^2\frac{\log^2 s}{n\sigma^2} + ...]^n = [1 + \frac{\log^2 s}{2n} + ...]^n$$

In the limit of large n,

$$\lim G_n(s) = e^{\frac{\log^2 s}{2}}$$

which we have already identified as the generating function for the normal distribution with mean 0 and variance 1.

5 Markov's inequality

This simple but useful result is on pp. 318-319 of the text.

Suppose that $h(x)$ is a function that assumes only non-negative values for all real x, like $|x|$, x^2, or x^4.

Consider the expectation of $h(X)$, and condition on whether X is or is not greater than some positive number c.

$$\mathbb{E}(h(X)) = \mathbb{P}(h(X) > c)\mathbb{E}(h(X)|(h(X) > c)) + \mathbb{P}(h(X) \leq c)\mathbb{E}(h(X)|(h(X) \leq c))$$

The first conditional expectation is no less than c and the second is not negative, so

$$\mathbb{E}(h(X)) \geq \mathbb{P}(h(X) > c) \cdot c$$

which is usually written as

$$\mathbb{P}(h(X) > c) \leq \frac{\mathbb{E}(h(X))}{c}$$

To get the most famous case, Chebyshov's inequality, choose a random variable X with $\mathbb{E}(X) = 0$ and some finite variance, choose $h(X) = X^2$, and set $c = a^2$. Then

$$\mathbb{P}(|X| > a) = \mathbb{P}(X^2 > a^2) \leq \frac{\mathbb{E}(X^2)}{a^2} = \frac{var(X)}{a^2}$$

.

This leads to another proof of the weak law of large numbers. When we sum up n independent copies of the random variable X,

$var(S_n) = n \cdot var(X)$

So by Chebyshov's inequality,

$$\mathbb{P}(|S_n| > a) \leq \frac{n \cdot var(X)}{a^2}$$

.

Now set $a = n\epsilon$ to obtain

$$\mathbb{P}(|S_n| > n\epsilon) \leq \frac{n \cdot var(X)}{(n\epsilon)^2}$$

.

$$\mathbb{P}(\frac{|S_n|}{n} > \epsilon) \leq \frac{var(X)}{n\epsilon^2}$$

.

This shows that for any fixed positive ϵ, the probability that the average of n independent copies of X exceeds ϵ goes to 0 at least as fast as $\frac{1}{n}$. But this is not entirely reassuring, since it leaves open the possibility that the average exceeds ϵ once for $10^6 < n < 10^7$, once for $10^7 < n < 10^8$, once for $10^8 < n < 10^9$, and so on. In other words, the average could still exceed ϵ for infinitely many values of n.

6 Strong law of large numbers

This is the classic proof, exercise 7.11.6 in the text. I think that some of the coefficients of terms that equal zero are incorrect in the book's solution.

Make the assumption (necessary for this proof, though not for the truth of the theorem) that $\mathbb{E}(X^4)$ has some finite value M.

Now take the expectation of

$$S_n^4 = (X_1 + X_2 + X_3 + ... + X_n)^4$$

.

This expansion contains a total of n^4 terms. Since all the X_i have the same distribution, the expectation of a term depends only on the exponents, not on the subscripts. There are five cases.

$\mathbb{E}(X_1^4) = \mathbb{E}(X_2^4) = ...$, and there are n such terms.

$\mathbb{E}(X_1^3 X_2) = \mathbb{E}(X_1^3 X_3) = \mathbb{E}(X_2^3 X_3) = ...$, and there are $4n(n-1)$ such terms.

$\mathbb{E}(X_1^2 X_2^2) = \mathbb{E}(X_1^2 X_3^2) = \mathbb{E}(X_2^2 X_3^2) = ...$, and there are $3n(n-1)$ such terms.

$\mathbb{E}(X_1^2 X_2 X_3) = \mathbb{E}(X_1^2 X_2 X_4) = \mathbb{E}(X_2^2 X_3 X_4) = ...$, and there are $6n(n-1)(n-2)$ such terms.

$\mathbb{E}(X_1 X_2 X_3 X_4) = \mathbb{E}(X_1 X_2 X_3 X_5) = \mathbb{E}(X_2 X_3 X_4 X_5) = ...$, and there are $n(n-1)(n-2)(n-3)$ such terms.

Check: $n+4n(n-1)+3n(n-1)+6n(n-1)(n-2)+n(n-1)(n-2)(n-3) = n^4$.

But all the X_i are independent, and so, for example,

$$\mathbb{E}(X_1^3 X_2) = \mathbb{E}(X_1^3)\mathbb{E}(X_2) = 0$$

because X_2 has zero mean. So only the terms with even exponents survive.

Thus $\mathbb{E}(S_n^4) = n\mathbb{E}(X_1^4) + 3n(n-1)\mathbb{E}(X_1^2)^2 = nM + 3n(n-1)var(X)^2$.

Now continue as in the previous proof.

By Markov's inequality, with $h(X) = X^4$

$$\mathbb{P}(|S_n| > a) = \mathbb{P}(S_n^4 > a^4) \leq \frac{\mathbb{E}(S_n^4)}{a^4}$$

.

Set $a = n\epsilon$ to obtain

$$\mathbb{P}(|S_n| > n\epsilon) \leq \frac{nM + 3n(n-1)var(X)^2}{(n\epsilon)^4}$$

.

$$\mathbb{P}(\frac{|S_n|}{n} > \epsilon) \leq \frac{C}{n^2\epsilon^4},$$

where C is some constant determined by M and $var(X)$.

This shows that for any fixed positive ϵ, the probability that the average of n independent copies of X exceeds ϵ goes to 0 at least as fast as $\frac{1}{n^2}$.

This is a much stronger and more reassuring result than before. Suppose you want to make sure that for $n \leq N$, the probability the the average of n copies of X will exceed ϵ even once is less than some small δ.

This is the probability of an infinite union, and it is no greater than the infinite sum of the probabilities of the individual events.

$$\mathbb{P}(\bigcup_{n=N}^{\infty}(|S_n| > n\epsilon)) \leq \sum_{n=N}^{\infty}\mathbb{P}(|S_n| > n\epsilon).$$

$$\mathbb{P}(\bigcup_{n=N}^{\infty}(|S_n| > n\epsilon)) \leq \sum_{n=N}^{\infty}\frac{C}{n^2\epsilon^4}.$$

So just choose N so large that

$$\sum_{n=N}^{\infty}\frac{C}{n^2\epsilon^4} < \delta$$

.

Another way of stating this conclusion is that the event $\frac{|S_n|}{n} > \epsilon$ cannot occur infinitely often, since you can choose N so large that the probability that it will ever occur again, even just once, is made arbitrary small. This is the content of the famous "first Borel-Cantelli lemma" to which the book's solution refers.