IEOR 3106: Introduction to Operations Research: Stochastic Models

Fall 2013, Professor Whitt, Thursday, September 12.

Sums of Independent Random Variables and Moment Generating Functions

1. §2.5. Sums of Independent Random Variables

(0) Dick and Jane meet at the Northwest Corner Building. A picture is worth a thousand formulas: Suppose that they plan to meet between 9 am and 10 am, but they will only wait for 10 minutes for the other. Suppose that they arrive independently at times that are uniformly distributed over the hour. What is the probability that they meet? Are they more likely to meet or not meet?

(i) What does it mean for two random variables X and Y to be **independent random variables**?

See Section 2.5.2. Pay attention to *for all*. We say that X and Y are independent random variables if

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y) \quad \text{for all} \quad x \quad \text{and} \quad y \ .$$

We can rewrite that in terms of cumulative distribution functions (cdf's) as We say that X and Y are independent random variables if

$$F_{X,Y}(x,y) \equiv P(X \le x, Y \le y) = F_X(x)F_Y(y) \quad \text{for all} \quad x \quad \text{and} \quad y \ .$$

When the random variables all have pdf's, that relation is equivalent to

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \text{for all} \quad x \quad \text{and} \quad y \ .$$

(ii) What is the *joint distribution* of (X,Y) in general?

See Section 2.5.

The joint distribution of X and Y is

$$F_{X,Y}(x,y) \equiv P(X \le x, Y \le y) \ .$$

(iii) How do we compute the distribution of the sum of two independent random variables?

The distribution of the sum of two independent random variables is the **convolution** of the distributions of the individual random variables. For example, if the two random variables have probability mass functions (pmf's) on the set of nonnegative integers, then

$$P(X+Y=k) = \sum_{j=0}^{k} P(X=j)P(Y=k-j) \tag{1}$$

for each integer k. There is a similar expression for probability density functions (pdf's) if the random variables have pdf's on the positive halfline. In particular,

$$f_{X+Y}(x)) = \int_0^x f_X(u)f_Y(x-u)\,du. \tag{2}$$

Example 2.36 (sum of two i.i.d. uniform random variables

Example 2.37 (sum of two independent Poisson random variables

2. §2.6. Moment Generating Functions

Given a random variable X, the *moment generating function* (mgf) of X (really of its probability distribution) is

$$\psi_X(t) \equiv E[e^{tX}] \ ,$$

which is a function of the real variable t, see Section 2.6 of Ross. (I here use ψ, whereas Ross uses ϕ.) An mgf is an example of a transform.

The random variable could have a continuous distribution or a discrete distribution;

Discrete case: Given a random variable X with a probability mass function (pmf)

$$p_n \equiv P(X = n), \quad n \geq 0, \ ,$$

the *moment generating function* (mgf) of X (really of its probability distribution) is

$$\psi_X(t) \equiv E[e^{tX}] \equiv \sum_{n=0}^{\infty} p_n e^{tn} \ .$$

The transform maps the pmf $\{p_n : n \geq 0\}$ (function of n) into the associated function of t.

Continuous case: Given a random variable X with a probability density function (pdf) $f \equiv f_X$ on the entire real line, the *moment generating function* (mgf) of X (really of its probability distribution) is

$$\psi(t) \equiv \psi_X(t) \equiv E[e^{tX}] \equiv \int_{-\infty}^{\infty} f(x) e^{tx} \, dx \ .$$

In the continuous case, the transform maps the pdf $\{f(x) : x \geq 0\}$ (function of x) into the associated function of t.

A major difficulty with the mgf is that it may be infinite or it may not be defined. For example, if X has a pdf $f(x) \equiv A/(1+x)^p, \quad x > 0$, for $p > 1$, then the mgf is infinite for all $t > 0$. Similarly, if X has the pmf $p(n) \equiv A/n^p$ for $n = 1, 2, \ldots$, then the mgf is infinite for all $t > 0$. As a consequence, probabilists often use other transforms. In particular, the characteristic function $E[e^{itX}]$, where $i \equiv \sqrt{-1}$, is designed to avoid this problem. We will not be using complex numbers in this class.

Two major uses of mgfs are: (i) calculating moments and (ii) characterizing the probability distributions of sums of random variables.

Below are some illustrative examples.

Examples 2.41 and 2.45: Poisson

Example 2.43 (2.42 in 9^{th} ed.) and 2.46: Normal

3. §2.8. Proofs of the LLN and the CLT: pp. 83-84 (pp. 82-83 in 9^{th} ed.)
We work with mgf's. We assume that each mgf $\psi_X(t) \equiv E[e^{tX}]$ is finite for some $t > 0$.
A key result behind these proofs is the *continuity theorem for mgf's.*

Theorem 0.1 (continuity theorem) *Suppose that X_n and X are real-valued random variables, $n \geq 1$. Let ψ_{X_n} and ψ_X be their mgf's (assumed finite). Then*

$$X_n \Rightarrow X \quad as \quad n \to \infty \quad (convergence\ in\ distribution)$$

if and only if

$$\psi_{X_n}(t) \to \psi_X(t) \quad as \quad n \to \infty \quad for\ all \quad t\ .$$

Now to prove versions of the law of large numbers (LLN) and the central limit theorem (CLT), we exploit the continuity theorem for mgf's and the following two lemmas:

Lemma 0.1 (convergence to an exponential) *If $\{c_n : n \geq 1\}$ is a sequence of real numbers such that $c_n \to c$ as $n \to \infty$, then*

$$(1 + (c_n/n))^n \to e^c \quad as \quad n \to \infty\ .$$

(This should be familiar when c_n is independent of n.)

The next lemma is classical Taylor series approximation applied to the mgf. For a function $h(t)$, recall that a Taylor series expansion is:

$$h(t) = h(0) + h'(0)t + h''(0)\frac{t^2}{2} \cdots$$

We also want a refinement with error term:

$$h(t) = \sum_{j=0}^{k} \frac{h^j(0)t^j}{j!} + o(t^k) \quad \text{as} \quad t \to 0,$$

where $o(t^k)$ is a quantity small compared to t^k, i.e., $o(t^k)/t^k \to 0$ as $t \to 0$.

Since the derivatives of the mgf evaluated at 0, are the moments, and the derivatives exist if the moments are finite, we get the following.

Lemma 0.2 (Taylor's theorem) *If $E[|X^k|] < \infty$, then the following version of Taylor's theorem is valid for the mgf $\psi_X(t) \equiv E[e^{tX}]$:*

$$\psi_X(t) = \sum_{j=0}^{j=k} \frac{E[X^j]t^j}{j!} + o(t^k) \quad as \quad t \to 0$$

where $o(t)$ is understood to be a quantity (function of t) such that

$$\frac{o(t)}{t} \to 0 \quad as \quad t \to 0\ .$$

Suppose that $\{X_n : n \geq 1\}$ is a sequence of IID random variables. Let

$$S_n \equiv X_1 + \cdots + X_n, \quad n \geq 1\ .$$

Theorem 0.2 (LLN) *If $E[|X|] < \infty$, then*

$$\frac{S_n}{n} \Rightarrow EX \quad as \quad n \to \infty\ .$$

Proof. Look at the mgf of S_n/n and then do a Taylor series approximation:

$$\psi_{S_n/n}(t) \equiv E[e^{tS_n/n}] = \psi_X(t/n)^n = (1 + \frac{tEX}{n} + o(t/n))^n$$

by the second lemma above. Hence, we can apply the first lemma to deduce that

$$\psi_{S_n/n}(t) \to e^{tEX} \quad \text{as} \quad n \to \infty.$$

By the continuity theorem for mgf's (convergence in distribution is equivalent to convergence of mgf's), the LLN is proved. ■

Theorem 0.3 (CLT) *If* $E[X^2] < \infty$, *then*

$$\frac{S_n - nEX}{\sqrt{n\sigma^2}} \Rightarrow N(0,1) \quad as \quad n \to \infty \ ,$$

where $\sigma^2 = Var(X) < \infty$.

Proof. For simplicity, consider the case of $EX = 0$. We get that case after subtracting the mean. Look at the mgf of $S_n/\sqrt{n\sigma^2}$:

$$\begin{aligned}
\psi_{S_n/\sqrt{n\sigma^2}}(t) &\equiv E[e^{t[S_n/\sqrt{n\sigma^2}]}] \\
&= \psi_X(t/\sqrt{n\sigma^2})^n \\
&= (1 + (\frac{t}{\sqrt{n\sigma^2}})EX + (\frac{t}{\sqrt{n\sigma^2}})^2\frac{EX^2}{2} + o(1/n))^n \\
&= (1 + \frac{t^2}{2n} + o(1/n))^n \\
&\to e^{t^2/2} = \psi_{N(0,1)}(t) \quad \text{as} \quad n \to \infty
\end{aligned}$$

by the two lemmas above. Thus, by the continuity theorem, the CLT is proved. ■
