Stat 230 Chapter 10

C.L.T., Normal Approximations and M.G.F.'s

2018/04/03

Abstract

This chapter covers two main topics, first one being the **Central Limit Theorem**, which says we can approximate the distribution of the sum and average of independent random variables that have a common mean and variance using a Normal distribution, and the second being **Moment Generating Functions**, which is a new set of tools that helps us analyze distributions.

Contents

1 Central Limit Theorem

1.1 Theorem

Theorem 1.1: Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ are *independent* (discrete or continuous) random variables all having the *same distribution*, with *common mean* μ and *variance* σ^2, then as $n \to \infty$,

1. the cumulative distribution function of the random variable
$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$
approaches the $N(0,1)$ cumulative distribution function;
2. the cumulative distribution function of
$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$
approaches the $N(0,1)$ cumulative distribution function.

When n is large (but finite), we will use C.L.T. to approximate the distribution of S_n or $\overline{X}$ by a Normal distribution. That is, we will use

$$S_n = \sum_{i=1}^n X_i \text{ has approximately a } N(n\mu, n\sigma^2) \text{ distribution for large } n$$

and

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \text{ has approximately a } N\left(\mu, \frac{\sigma^2}{n}\right) \text{ distribution for large } n.$$

1.2 Remarks

1. C.L.T. holds if and only if the common mean and variance exist.
2. C.L.T. approximates the distribution of the sum (S_n) or average ($\overline{X}$).
   - The accuracy of this approximation depends on n (bigger is better) and the actual distribution of the X_i's (more symmetric is better).
   - If the X_i's themselves have a Normal distribution, then S_n and $\overline{X}$ as linear combinations of Normal distributions will have exactly Normal distributions for all values of n (see Chapter 9).
   - The more non-Normal-shaped X_i's are, the larger n we need in order to make a good approximation.

1.3 Problem Solving Strategies

1.3.1 When $X_i \sim N(\mu, \sigma^2)$

Given $n = 9, \mu = 0.128, \sigma = 0.005$, find the probability of $S_n \geq 1$.

- **Define unknown**: $S_8 = \sum_{i=1}^{8} X_i$.
- **Define distribution**: By C.L.T., $S_8 \sim N(8\mu, 8\sigma^2) = N(1.024, 0.0002)$.
- **Calculate probability**:

$$\begin{aligned} P(S_8 \geq 1) &\approx P\left(Z \geq \frac{1 - 1.024}{\sqrt{0.0002}}\right) \\ &= P(Z \geq -1.70) \\ &= 0.95543. \end{aligned}$$

1.3.2 When $X_i \sim Exponential(\theta)$

Given $\lambda = 6$ (or $\theta = 1/6$), find $P(83 < S_{500} \leq 84)$.

- **Define unknown**: $S_{500} = \sum_{i=1}^{500} X_i$.
- **Define distribution**: By C.L.T., $S_{500} \sim N(500\mu, 500\sigma^2)$ where $\mu = E(X_i)$ and $\sigma^2 = Var(X_i)$. For the exponential distribution, $\mu = \theta = 1/6$, $\sigma^2 = \theta^2 = 1/36$, so $S_{500} \sim N\left(\frac{500}{6}, \frac{500}{36}\right)$.
- **Calculate probability**:

$$\begin{aligned} P(83 < S_{500} \leq 84) &\approx P\left(\frac{83 - \frac{500}{6}}{\sqrt{\frac{500}{36}}} < Z \leq \frac{84 - \frac{500}{6}}{\sqrt{\frac{500}{36}}}\right) \\ &= P(-0.09 < Z \leq 0.18) \\ &= 0.10728. \end{aligned}$$

1.3.3 When $X_i \sim Poisson(\mu)$

Theorem 1.2: Normal Approximation to Poisson

Suppose $X \sim Poisson(\mu)$. Then the cumulative distribution function of the standardized random variable

$$Z = \frac{X - \mu}{\sqrt{\mu}}$$

approaches that of a standard Normal random as $\mu \to \infty$.

1.3.4 When $X_i \sim Binomial(n, p)$

Theorem 1.3: Normal Approximation to Binomial

Suppose $X \sim Binomial(n, p)$. Then for n large, the random variable

$$W = \frac{X - np}{\sqrt{np(1-p)}}$$

has approximately a $N(0, 1)$ distribution. In other words,

$$\frac{X - np}{\sqrt{np(1-p)}} \sim N(0, 1) \text{ and } X \sim N(np, np(1-p))$$

1.3.5 Notes on Continuity Correction

Remark 1: Continuity Correction

- A continuity correction should **NOT** be applied when approximating a continuous distribution by the Normal distribution.
- Whenever approximating the probability of a single value for a discrete distribution, such as $P(X = 50)$ where $X \sim Binomial(100, 0.5)$, you **NEED** to use the continuity correction. Otherwise, for approximating the Binomial with large n, it is not necessary to use the correction.
- Rather than trying to guess or memorize when to add or subtract 0.5, it is often helpful to sketch a histogram and shade the bars you wish to include.
- **Rule of thumb**: use the continuity correction whenever the standard deviation of the integer-valued random variable begin approximated is less than 10.

2 Moment Generating Functions

2.1 Introduction to M.G.F.

Definition 2.1: M.G.F - Moment Generating Function

Consider a discrete random variable X with probability function $f(x)$. The **moment generating function** of X is defined as

$$M(t) = E(e^{tX}) = \sum_{\forall x} e^{tx} f(x).$$

We will assume that for some $a > 0$, $\sum_x e^{tx} f(x) < \infty$ for all $t \in [-a, a]$.

Theorem 2.1: Generating Moments

Let the random variable X have m.g.f. $M(t)$. Then

$$E(X^k) = M^{(k)}(0)$$

for $k = 1, 2, \ldots$ where $M^{(k)}(0)$ stands for the kth derivative of $M(t)$ evaluated at $t = 0$.

Proof: Suppose the sum $M(t) = \sum_{\forall x} e^{tx} f(x)$ converges, then

$$\begin{aligned} M^{(k)}(t) &= \frac{d}{dt^k} \sum_{\forall x} e^{tx} f(x) \\ &= \sum_{\forall x} \frac{d}{dt^k} (e^{tx}) f(x) \\ &= \sum_{\forall x} x^k (e^{tx}) f(x) \end{aligned}$$

Therefore $M^{(k)}(0) = \sum_x x^k f(x) = E(X^k)$ as stated. ■

Remark 2: Moments

In statistics, a moment is a specific quantitative measure of the shape of a set of points. When the points represent probability density, the zeroth moment is the **total probability** (i.e. one); the first moment is the **expectation value** , the second moment is the **variance**, the third moment of the skewness, and the fourth moment is the **kurtosis**.

2.2 M.G.F. for the Binomial Distribution

Theorem 2.2: M.G.F for $X \sim Binomial(n,p)$

Let $X \sim Binomial(n,p)$.

$$\begin{aligned} M(t) &= \sum_{\forall x} e^{tx} f(x) \\ &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \qquad \Rightarrow x = (pe^t), y = (1-p), r = x \\ &= (pe^t + 1 - p)^n \end{aligned}$$

Therefore

$$\begin{aligned} M'(t) &= npe^t(pe^t + 1 - p)^{n-1} \\ &\Rightarrow E(X) = M'(0) = np \end{aligned}$$

$$\begin{aligned} M''(t) &= npe^t(pe^t + 1 - p)^{n-1} + n(n-1)p^2 e^{2t}(pe^t + 1 - p)^{n-2} \\ &\Rightarrow E(X^2) = M''(0) = np + n(n-1)p^2 \\ &\Rightarrow Var(X) = E(X^2) - E(X)^2 = np(1-p) \end{aligned}$$

Remark 3: Binomial Theorem Revisit

Let $n \in \mathbb{Z}^+; x, y \in \mathbb{F}$. The coefficient of $x^k y^{n-k}$ in the kth term in the expansion of $(x+y)^n$, is equal to

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$

The below process is known as **binomial expansion**:

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^{n-r} y^r = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}$$

Note that, in the binomial coefficient, if we ignore the factorials, the sum of the denominator is equal to the sum of numerator. This may help you manipulate certain algebra quantities into a binomial form.

.

2.3 M.G.F. for the Poisson Distribution

Theorem 2.3: M.G.F for $X \sim Poisson(\mu)$

Let $X \sim Poisson(\mu)$.

$$\begin{aligned} M(t) &= \sum_{\forall x} e^{tx} f(x) \\ &= \sum_{x=0}^{\infty} e^{tx} \frac{\mu^x}{x!} e^{-\mu} \\ &= e^{-\mu} \sum_{x=0}^{\infty} \frac{(e^t \mu)^x}{x!} \\ &= e^{-\mu} e^{e^t \mu} \end{aligned}$$

Therefore $\forall t \in \mathbb{R}$

$$\begin{aligned} M'(t) &= \mu e^t \exp(-\mu + \mu e^t) \\ &\Rightarrow E(X) = M'(0) = \mu \end{aligned}$$

$$\begin{aligned} M''(t) &= \exp(-\mu + \mu e^t)\mu^2 e^{2t} + \exp(-\mu + \mu e^t)\mu e^t \\ &\Rightarrow E(X^2) = M''(0) = \mu^2 + \mu \\ &\Rightarrow Var(X) = E(X^2) - (E(X))^2 = \mu^2 + \mu - \mu^2 = \mu. \end{aligned}$$

Remark 4: Exponential Function as Power Series Revisit

Most commonly, the exponential function is formally defined by the following power series:

$$\exp(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots$$

The constant e can thus be defined as

$$e = \exp(1) = \sum_{k=0}^{\infty} \frac{1}{k!}.$$

The exponential function can also be defined in the following two ways:

1. The solution y to the equation
$$x = \int_1^y \frac{1}{t} dt$$

2. The limit
$$e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$$

2.4 Uniqueness of M.G.F.

The moment generating function uniquely identifies a distribution in the sense that, if two random variables have the same moment generating function (agree as functions, not just a few points), they have the same distribution (i.e. the same *pdf*, *cdf*, moments, etc). This means that if we were able to determine the m.g.f. for a given random variable, then it can be used to identify the distribution. This gives us another technique for finding the distribution of a random variable.

Theorem 2.4: Uniqueness Theorem of Moment Generating Functions

Suppose that random variables X and Y have moment generating functions $M_X(t)$ and $M_Y(t)$ respectively. If $M_X(t) = M_Y(t)$ for all t then X and Y have the same distribution.

2.5 Convergence of M.G.F.

Moment generating functions can also be used to determine that a sequence of distributions gets closer and closer to some limiting distribution.

Theorem 2.5: Convergence Theorem of Moment Generating Functions

Suppose X_n has mgf $M_n(t)$ and $M_n(t) \to M(t)$ for each t such that $M(t) < \infty$. If $M(t)$ is the mgf of X, then X_n converges in distribution to X.

Proof: Suppose that a sequence of probability functions $f_n(x)$ have corresponding moment generating functions

$$M_n(t) = \sum_{\forall x} e^{tx} f(x).$$

Suppose moreover that $f_n(x) \to f(x)$ point-wise in x as $n \to \infty$. Then since

$$f_n(x) \to f(x) \text{ as } n \to \infty \text{ for each } x,$$

$$\sum_{\forall x} e^{tx} f_n(x) \to \sum_{\forall x} e^{tx} f(x) \text{ as } n \to \infty \text{ for each } t.$$

which says that $M_t(t)$ converges to $M(t)$, the mgf of the limiting distribution. ■

2.6 M.G.F. of a Continuous Random Variable

Definition 2.2: M.G.F. of a Continuous Random Variable

Consider a continuous random variable X with *pdf* $f(x)$. The *mgf* of X is defined as

$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx.$$

We assume that for some $a > 0$, $\int_{-\infty}^{\infty} e^{tx} f(x)dx < \infty$ for all $t \in [-a, a]$.

2.7 M.G.F. of Continuous Distributions

Theorem 2.6: M.G.F. of the Normal distribution

$$M(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2}).$$

Theorem 2.7: M.G.F. of the Exponential distribution

$$M(t) = \frac{1}{1 - \theta t} \text{ for } t < \frac{1}{\theta}.$$