when it's almost course evaluation time & my teacher starts trying me

Today's Agenda

Last time:

- Linear combinations of independent normal random variables
- Indicator random variables

Today (Lec 35, 07/22):

- Chapter 10!!

Recall: If $X_1, ..., X_n$ are independent with $X_i \sim N(\mu, \sigma^2)$, then

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Question: What if $X_1, ..., X_n$ are not normally distributed? Then what is the **distribution** of the sample mean?

Theorem (Central Limit Theorem)

Suppose that $X_1, ..., X_n$ are independent random variables, each with a common cumulative distribution function F. Suppose further that $E(X_i) = \mu$, and $Var(X_i) = \sigma^2 < \infty$.
Then for all $x \in \mathbb{R}$

$$P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq x\right) \to \Phi(x),$$

as $n \to \infty$.

In other words, if n is large

$$\bar{X} \overset{approx}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{and} \quad \sum_{i=1}^{n} X_i \overset{approx}{\sim} N\left(n\mu, n\sigma^2\right)$$

Example

Harold is eating a box of chocolate right now (yes, right now). Each box contains 20 cubes, and it is supposed to have 500 grams of chocolate in it. The weight of each chocolate cube varies a little because they are hand-made from Switzerland. The weight W of each cube is a random variable with mean $\mu = 25$ grams, and the standard deviation $\sigma = 0.1$ grams. Find the probability that a box has at least 500 grams of chocolate in it, assuming that the weight of each cube is independent.

The point of this example is that, as long as you:

- have a set of independent and identically distributed random variables,
- have finite common mean μ and finite common variance,

the distribution of their sample mean can be approximated by normal distribution.

If the random variables are normal, then the sample mean is **exactly** normal.

If they aren't then their sample mean is **approximately** normal.

Example

Jason rolls a six sided die 1000 times, and records the results. If the die is a fair die, estimate the probability that the sum of the die rolls is less than 3400.

Theorem

If $X_n \sim Binomial(n, p)$, then for large n

$$\frac{X_n - np}{\sqrt{np(1-p)}} \overset{approx}{\sim} N(0,1)$$

Theorem

If $X_\lambda \sim Poi(\lambda)$, then for large λ

$$\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \overset{approx}{\sim} N(0,1)$$

BEWARE: When approximating discrete random variables with normal distribution, that discrete distribution will never be truly “normal”.

That is because countable infinity is still “smaller” than a continuous interval with respect to the “number” of numbers.

So, in this case, we use what’s called continuity correction.

How to correct for continuity

Let's say we want to approximate the probability of some discrete random variable S_n with normal distribution.

$$
\begin{aligned}
P(S_n = s) &= P(s - 0.5 < S_n < s + 0.5) \\
&\approx P\left(\frac{(s - 0.5) - \mu_{S_n}}{SD(S_n)} < Z < \frac{(s + 0.5) - \mu_{S_n}}{SD(S_n)} \right)
\end{aligned}
$$

Example

Suppose $X \sim Poisson(\lambda)$. Use normal approximation to estimate $P(X > \lambda)$. Compare this approximation with the true value when $\lambda = 9$.

Example

Suppose that Jason flips a fair coin 1000 times. Approximate the probability that the number of heads is between 450 and 550.

Rules of thumb for using the central limit theorem:

- In general if the number of observations exceeds 30, then the central limit theorem often provides a reasonable approximation
- If the distribution of the observations is "close" to being unimodal, not too skewed, and is "close" to being continuous, then the central limit approximation can be reasonable for even smaller values of n (5-15).
- If the distribution is highly skewed, or very discrete, then a larger value of n might be necessary: $(n > 50)$
- When approximating a **continuous** distribution with normal, we do not use continuity correction.

Moment generating functions

So far, we have two functions that can define/characterise the distribution of a RV:

a) probability function/probability density function $f(x)$,

b) cumulative distribution function $F(x)$.

Another one will be the moment generating function!

Definition

The **Moment generating function** or MGF of a random variable X is given by

$$M_X(t) = \mathrm{E}[e^{tX}], \quad t \in \mathbb{R}$$

In particular, if X is discrete with p.f. $f(x)$ then

$$M_X(t) = \sum_{x \in X(S)} e^{tx} f(x), \quad t \in \mathbb{R}.$$

Though we won't cover it in this course, if X is continuous,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx, \quad t \in \mathbb{R}.$$

Recall that integral over $(-\infty, \infty)$ implicitly means "over the range of X".

Properties of the mgf

a)

$$M_X(0) = 1$$

b)

$$M_X(t) = \sum_{j=0}^{\infty} \frac{t^j \mathrm{E}[X^j]}{j!}$$

c) So long as $M_X(t)$ is defined in a neighbourhood of $t = 0$,

$$\frac{d}{dt^k} M_X(0) = \mathrm{E}[X^k]$$

The significance of MGF is that, under certain conditions, it can show equivalence of two distributions.

Theorem (Continuity theorem)

If X and Y have MGFs $M_X(t)$ and $M_Y(t)$ defined in neighbourhoods of the origin, and satisfying $M_X(t) = M_Y(t)$ for all t where they are defined, then

$$X \stackrel{D}{=} Y.$$

This means that the MGF uniquely characterizes a distribution.

Example

Let $X \sim Poi(\lambda)$. Derive the MGF of X, and use it to show that

$$\mathrm{E}[X] = \lambda = Var(X).$$