Mgf of a Normal (μ, σ^2)

$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \quad \text{for} \quad -\infty < x < \infty.$$

$$M(s) = E(e^{sX}) = \int_{-\infty}^{\infty} e^{sx} \cdot \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sx} \, e^{-\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2)} \, dx.$$

$$= \frac{e^{-\frac{\mu^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left\{x^2 - 2(\mu + \sigma^2 s)x\right\}} \, dx.$$

Let's complete the square in the term between the Curly brace { } in the exponent ...

$$x^2 - 2(\mu + \sigma^2 s)x + (\mu + \sigma^2 s)^2 - (\mu + \sigma^2 s)^2$$

$$= (x - (\mu + \sigma^2 s))^2 - (\mu + \sigma^2 s)^2$$

then substituting this back into the exponent ...

$$= \frac{e^{-\frac{\mu^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left\{(x-(\mu+\sigma^2 s))^2 - (\mu+\sigma^2 s)^2\right\}} dx$$

$$= \frac{e^{-\frac{\mu^2}{2\sigma^2} + \frac{(\mu+\sigma^2 s)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \underbrace{\boxed{\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left\{\frac{x-(\mu+\sigma^2 s)}{\sigma}\right\}^2} dx}}_{\sigma\sqrt{2\pi}}$$

$$= e^{\frac{(\mu+\sigma^2 s)^2 - \mu^2}{2\sigma^2}} = e^{\frac{\cancel{\mu^2} + 2\mu\sigma^2 s + \sigma^4 s^2 \cancel{-\mu^2}}{2\sigma^2}}$$

$$\boxed{M(s) = e^{\mu s + \frac{\sigma^2 s^2}{2}}.} \quad \text{(whew!)}$$

the mgf of a Normal(μ, σ^2).

$\blacksquare$

<u>Corollary</u> The mgf of a Standard normal is

$$M_Z(s) = e^{\frac{s^2}{2}}. \quad \text{for all real } s.$$

Remark

Suppose $X \sim \text{Normal}(\mu, \sigma^2)$.

Let's find the distribution of $Z = \dfrac{X-\mu}{\sigma}$ using mgfs.

$$M_Z(s) = E(e^{sZ}) = E\left(e^{s\left(\frac{X-\mu}{\sigma}\right)}\right)$$

$$= E\left(e^{\left(\frac{s}{\sigma}\right)X} e^{-\frac{\mu s}{\sigma}}\right)$$

$$= e^{-\frac{\mu s}{\sigma}} E\left(e^{\left(\frac{s}{\sigma}\right)X}\right)$$

$$= e^{-\frac{\mu s}{\sigma}} M_X\left(\frac{s}{\sigma}\right) = e^{-\frac{\mu}{s}} e^{\mu\left(\frac{s}{\sigma}\right) + \frac{\sigma^2\left(\frac{s}{\sigma}\right)^2}{2}}$$

$$= e^{-\frac{\mu s}{\sigma}} e^{\frac{\mu s}{\sigma} + \frac{s^2}{2}} = e^{\frac{s^2}{2}}.$$

which is the mgf of a standard Normal.

Therefore if $X \sim \text{Normal}(\mu, \sigma^2)$ then

$$Z = \frac{X-\mu}{\sigma} \sim \text{Normal}(0,1).$$

and we have another proof of this important fact.

Another interesting example

Suppose $X_1, X_2, \ldots, X_n$ are independent Normally-distributed r.v.s that all have the same μ and σ^2 parameters.

In Statistics, we usually write this as

$$X_1, X_2, \ldots, X_n \sim \text{iid Normal}(\mu, \sigma^2)$$

where iid means

Independent and Identically Distributed.

Consider the r.v.

$$\overline{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Let's compute the mgf of $\overline{X}_n$.

$$M_{\overline{X}_n}(s) = E(e^{s\overline{X}_n}) = E\left(e^{s\left(\frac{X_1 + \cdots + X_n}{n}\right)}\right) = E\left(e^{\left(\frac{s}{n}\right)\{X_1 + \cdots + X_n\}}\right)$$

$$= E\left(e^{\left(\frac{s}{n}\right)X_1}\right) E\left(e^{\left(\frac{s}{n}\right)X_2}\right) \cdots E\left(e^{\left(\frac{s}{n}\right)X_n}\right) \quad \text{by independence}$$

$$= M_{X_1}\left(\tfrac{s}{n}\right) M_{X_2}\left(\tfrac{s}{n}\right) \cdots M_{X_n}\left(\tfrac{s}{n}\right) = \left(e^{\mu\left(\frac{s}{n}\right) + \frac{\sigma^2 (s/n)^2}{2}}\right)^n = e^{\mu s + \frac{\sigma^2 s^2}{2n}}$$

That is, $\bar{X}_n$ has the moment-generating function of a Normal r.v. with mean μ and variance σ^2/n. So, it must follow that

$$\bar{X}_n \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right).$$

An easy exercise is now show:

if $X_1, \ldots, X_n$ are independent

$X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$

then $$X_1 + X_2 + \cdots + X_n \sim \text{Normal}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Solution:

$M_{X_1+\cdots+X_n}(s) = \prod_{i=1}^n M_{X_i}(s)$ by independence.

$$= \prod_{i=1}^n e^{\mu_i s + \frac{\sigma_i^2 s^2}{2}} = e^{\left(\sum_{i=1}^n \mu_i\right)s + \frac{\left(\sum_{i=1}^n \sigma_i^2\right)s^2}{2}}$$

which is the mgf of a $\text{Normal}\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$.

$\blacksquare$

The following result is of central importance to all of Statistics.

The Central Limit theorem

Suppose $X_1, X_2, \ldots, X_n \sim$ i.i.d (not necessarily Normally-distributed) but the common distribution of these r.v.s has mean $= \mu$ and variance $= \sigma^2$.

Then for all z,

$$\lim_{n \to \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le z \right) = \Phi(z).$$

or, equivalently,

$$\lim_{n \to \infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le z \right) = \Phi(z).$$

<u>Remark</u> A large sum of small independent r.v.s. tends to have a Normal distribution.

<u>Remark</u> The result above is not only theoretically important but also Practically Important. For example if n is "Large"

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \approx Z \text{ a standard Normal r.v.}$$

Example Roll a balanced die 100 times.

Estimate the probability that their sum is ≤ 300

If we let X_i be the result on roll i, then

$X_1, X_2, \ldots, X_{100}$ are independent and

all have mean $\mu = \frac{7}{2} = 3.5$ and $\sigma^2 = \frac{35}{12} = 2.916\bar{6}$

(or $\sigma \doteq 1.707825$). Therefore by the Central

Limit theorem:

$$P\left(\sum_{i=1}^{100} X_i \leq 300\right) = P\left(\frac{\sum_{i=1}^{100} X_i - 350}{(1.707825)(10)} \leq \frac{300-350}{17.07825}\right)$$

$$\approx P(Z \leq -2.93) = 1 - P(Z > -2.93) =$$
$$= 1 - .9983 = .0017.$$

(less than a .2% chance $\leftarrow$ very rare!)

The proof of the Central Limit theorem

We assume that the identical distribution possesses a moment-generating function — this will simply the proof greatly. (But it should be noted the statement of the Central limit theorem does not need this condition and generally the conclusion of the CLT holds under much weaker assumptions).

Suppose $X_1, X_2, \ldots, X_n \sim$ iid and that $M(s)$ is the common mgf. Then we consider the r.v.

$$Z_n := \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}, \quad \text{where}$$

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$ is the sample mean.

The strategy of the proof is to show the mgf of Z_n;

$M_n(s) = E\left(e^{sZ_n}\right)$, converges to $e^{\frac{s^2}{2}}$ (the mgf of a standard normal) by showing

$$\ln M_n(s) \longrightarrow \frac{s^2}{2} \quad \text{as} \quad n \to \infty.$$

Since the mgf, when it exists, uniquely identifies the distribution, we must have the limit distribution of Z_n is converging to a standard Normal.

$$E(e^{sZ_n}) = E\left(e^{s\left(\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\right)}\right) = E\left(e^{\frac{s\sqrt{n}}{\sigma}\bar{X}_n}\, e^{-\frac{\mu s\sqrt{n}}{\sigma}}\right)$$

$$= e^{-\frac{\mu s\sqrt{n}}{\sigma}}\, E\left(e^{\frac{s\sqrt{n}}{\sigma n}(X_1+\cdots+X_n)}\right) = e^{-\frac{\mu s\sqrt{n}}{\sigma}}\left\{E\left(e^{\frac{s}{\sigma\sqrt{n}}X_1}\right)\right\}^n.$$

$$= e^{-\frac{\mu s\sqrt{n}}{\sigma}}\left\{M_{X_1}\left(\tfrac{s}{\sigma\sqrt{n}}\right)\right\}^n$$

$$= e^{-\frac{\mu s\sqrt{n}}{\sigma}}\left\{1 + \frac{s}{\sigma\sqrt{n}}\mu + \frac{s^2}{2\sigma^2 n}(\sigma^2+\mu^2) + \overbrace{O(n^{-3/2})}^{\text{terms that have } n^{3/2} \text{ or higher powers of } n \text{ in the denominator.}}\right\}^n$$

Now take natural logarithms on both sides:

$$\ln E(e^{sZ_n}) = -\frac{\mu s\sqrt{n}}{\sigma} + n\ln\Big(1 + \overbrace{\Big[\frac{s\mu}{\sigma\sqrt{n}} + \frac{s^2(\sigma^2+\mu^2)}{2\sigma^2 n} + O(n^{-3/2})\Big]}^{x}\Big)$$

recall that the taylor expansion of $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots$
to get

$$\ln E(e^{sZ_n}) = -\frac{\mu s\sqrt{n}}{\sigma} + n\left\{\left[\frac{s\mu}{\sigma\sqrt{n}} + \frac{s^2(\sigma^2+\mu^2)}{2\sigma^2 n} + o(n^{-3/2})\right] - \frac{1}{2}\left[\frac{s\mu}{\sigma\sqrt{n}} + \frac{s^2(\sigma^2+\mu^2)}{2\sigma^2 n} + O(n^{-3/2})\right]^2 + O(n^{-3/2})\right\}$$

$$= -\frac{\mu s\sqrt{n}}{\sigma} + n\left\{\frac{s\mu}{\sigma\sqrt{n}} + \frac{s^2(\sigma^2+\mu^2)}{2\sigma^2 n} + O(n^{-3/2}) - \frac{1}{2}\left[\frac{s^2\mu^2}{\sigma^2 n} + o(n^{-3/2})\right] + O(n^{-3/2})\right\}$$

$$= \cancel{-\frac{\mu s\sqrt{n}}{\sigma}} + \cancel{\frac{s\mu\sqrt{n}}{\sigma}} + \frac{s^2(\sigma^2+\mu^2)}{2\sigma^2} - \frac{s^2\mu^2}{2\sigma^2} + O(n^{-1/2})$$

$= \frac{s^2}{2} + O(n^{-1/2})$ ← as $n\to\infty$ this term will go to zero. Therefore,

$$\lim_{n\to\infty}\ln E(e^{sZ_n}) = \frac{s^2}{2} \implies E(e^{sZ_n}) \to e^{s^2/2} \text{ and } Z_n \to \text{a standard normal.}$$

∎

Example Suppose $X_1, X_2, \dots, X_{100}$ are the lifetimes of 100 (working) independent lightbulbs, i.e., each X_i has pdf

$$f(x) = \begin{cases} \frac{1}{10000} e^{-\frac{x}{10000}} & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases}$$

We saw that $E(X_i) = 10000$ (hours), while the median lifetime, i.e., the value of m such that

$$P(X_i \le m) = \frac{1}{2} \iff \int_0^m \frac{1}{10000} e^{-\frac{x}{10000}} dx = 1 - e^{-\frac{m}{10000}} = 1/2$$

$$\iff e^{-\frac{m}{10000}} = 1/2 \iff m \doteq 6931.472 \text{ (hours)}.$$

Estimate the probability that the average (sample mean) of our 100 lightbulbs will less than or equal to 7000 hours.

i.e. Estimate

$$P(\overline{X}_{100} \le 7000).$$

By the Central Limit theorem $\frac{\overline{X}_{100} - \mu}{\sigma/\sqrt{100}} \approx$ a standard Normal

where $\mu = \underbrace{10000}_{10^4}$ and $\sigma^2 = 10^8 \Rightarrow \sigma = 10^4 \Rightarrow \frac{\sigma}{\sqrt{100}} = 10^3 = 1000.$

$$P\left(\frac{\overline{X}_{100} - \mu}{\frac{\sigma}{\sqrt{100}}} \le \frac{7000 - 10{,}000}{1000} \right)$$

$$\approx P(Z \le -3) \doteq .0013.$$

~~How large should n be?~~

The Central Limit theorem is still useful when the size of the sample is small but requires the "population" distribution to be symmetric about its mean.

For example, if $X_1, X_2, X_3, \ldots, X_n \sim$ iid uniform$(0,1)$ then $\overline{X}_n$ will be well-approximated by a Normal distribution for value of n as small as 5 or 6.

But generally for values of $n \approx 10$ or more when population has symmetry the Central Limit theorem will give good approximations.

If we do not have symmetry, then usually $n \geq 30$ will suffice, but this is not generally accepted. If the r.v. is bounded then the CLT with $n \approx 30$ is generally believed to give good results.

Remark

Recall that the sum S_n of n independent Gamma(1, 1) r.v.s has a Gamma(n, 1) distribution; i.e. has pdf

$$f(x) = \frac{x^{n-1} e^{-x}}{(n-1)!} \quad \text{for } x > 0.$$

The mean of this distribution is $\alpha\beta = n(1) = n$.

and variance is $\alpha\beta^2 = n(1)^2 = n$ so the standard deviation is $\sigma = \sqrt{n}$. The pdf of $Z_n := \frac{S_n - n}{\sqrt{n}}$ is found by the cdf method:

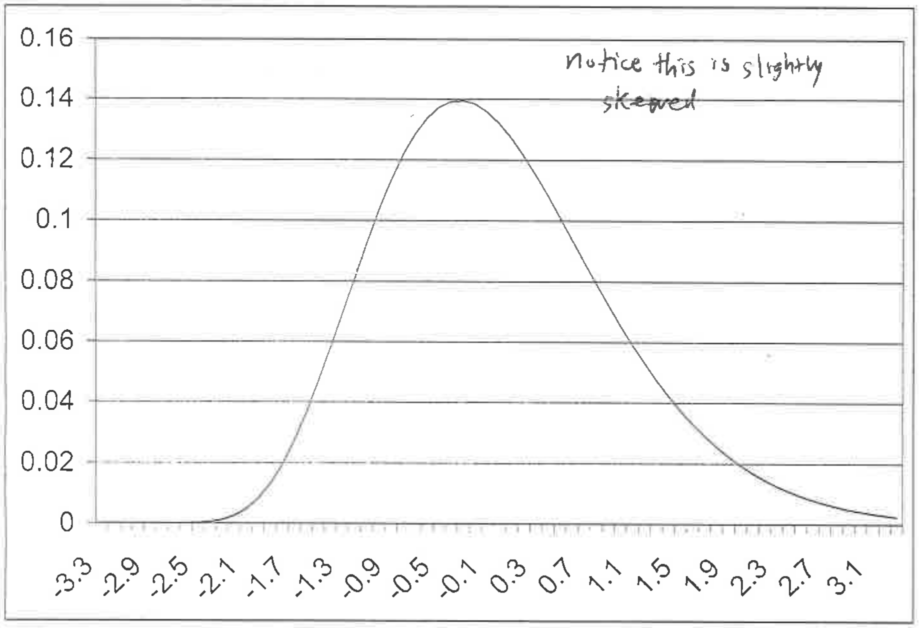
$$P(Z_n \le z) = P\left(\frac{S_n - n}{\sqrt{n}} \le z\right) = P\left(S_n \le n + z\sqrt{n}\right)$$

$$= \int_0^{n+z\sqrt{n}} \frac{x^{n-1} e^{-x}}{(n-1)!}\, dx.$$

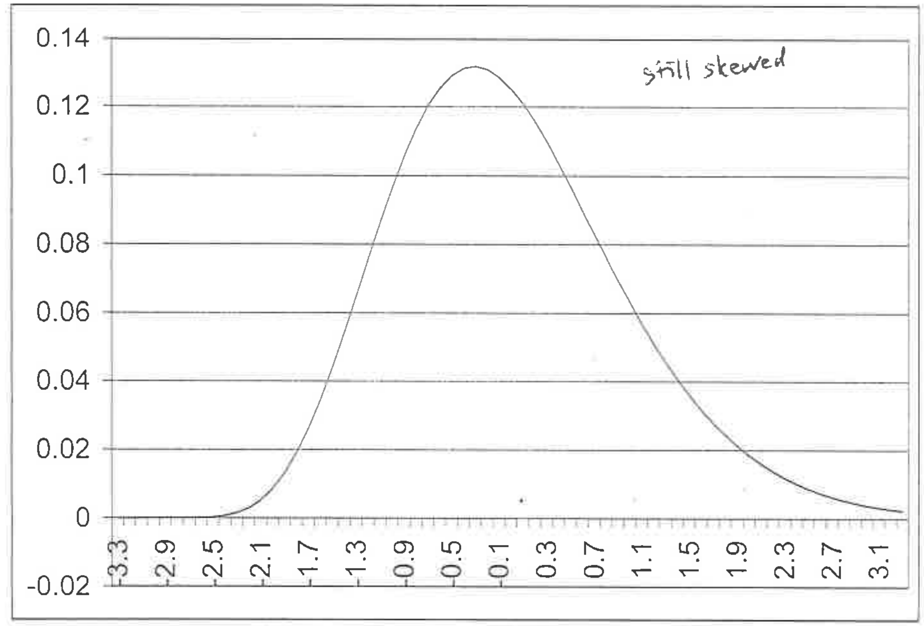
$$f_{Z_n}(z) = \frac{(n+z\sqrt{n})^{n-1} e^{-(n+z\sqrt{n})}}{(n-1)!} \cdot \sqrt{n} \quad \text{for } -n \le z < \infty.$$

Some Plots of this pdf for $n = 9, 10, 30, 100$ follow:

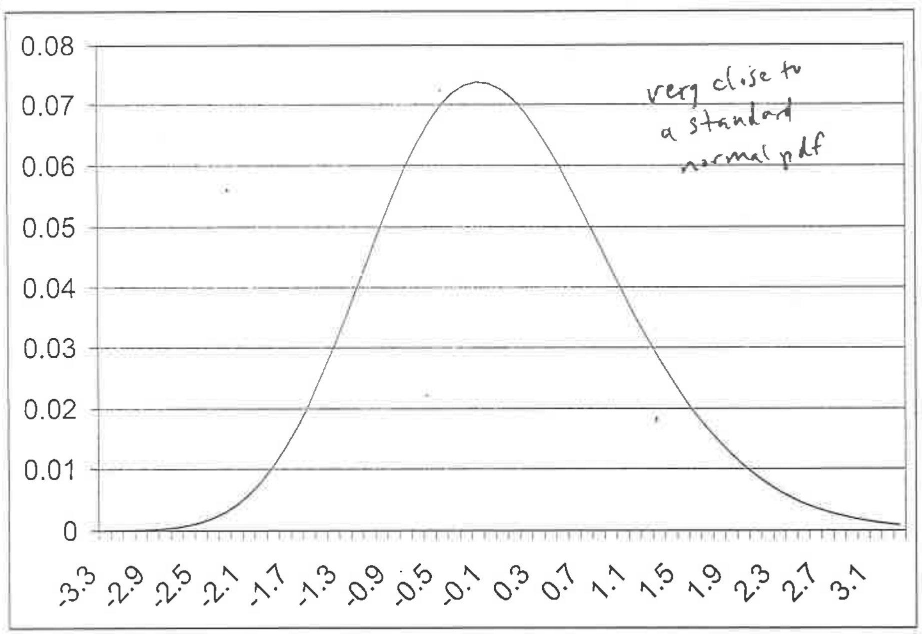
$n=9$



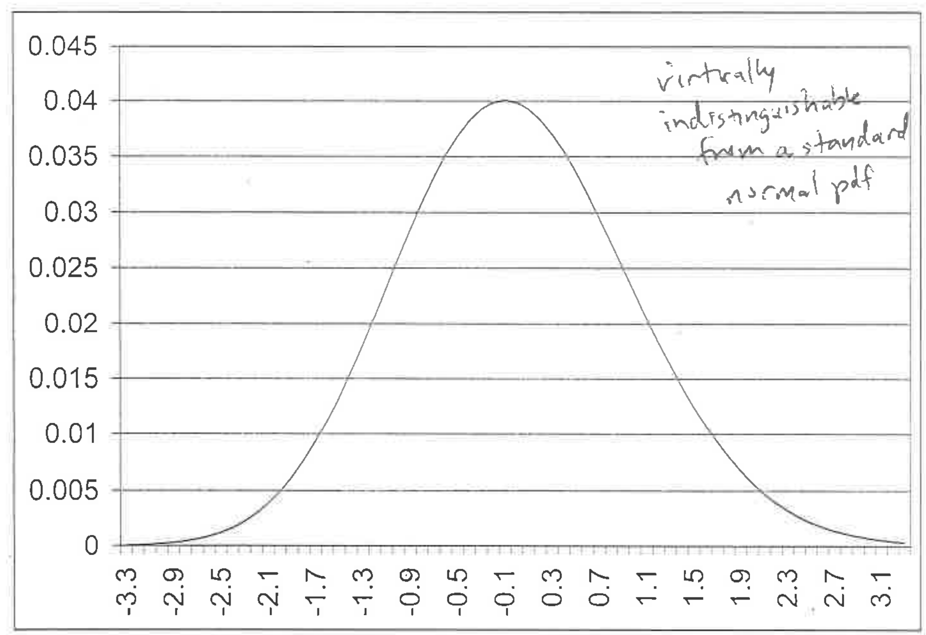
$n=10$



$n=30$



$n=100$



Using CLT on discrete populations....

When using the CLT ~~on~~ populations that are discrete sometimes the approximation can be improved by a <u>Continuity correction</u>.

For example, if $X_1, X_2, \ldots, X_{20} \sim$ i.i.d. Bernoulli$(\frac{1}{2})$

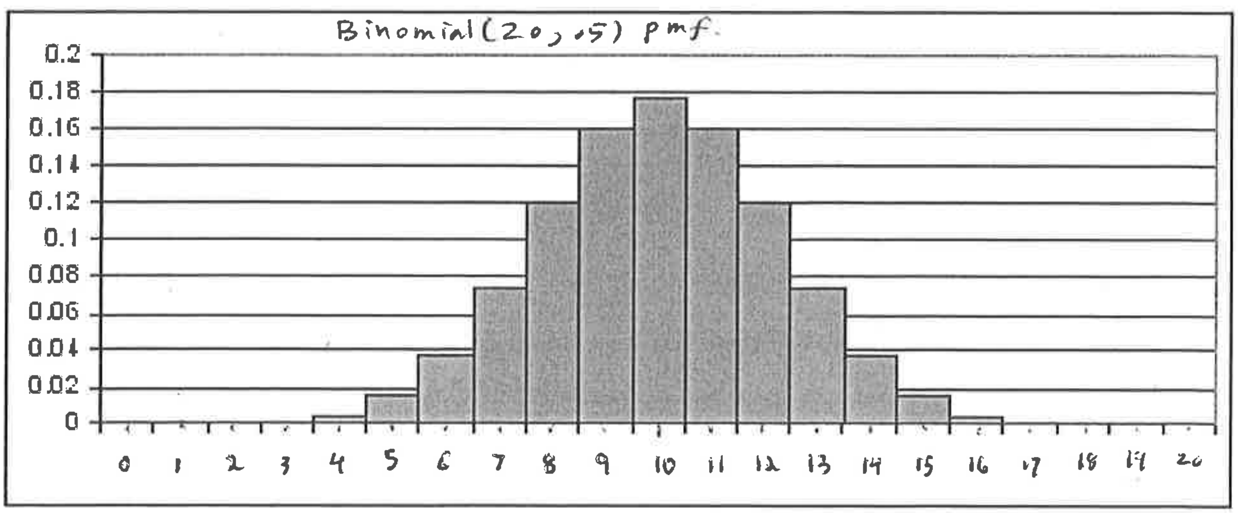
then $\dfrac{\overline{X}_{20} - p}{\frac{p(1-p)}{\sqrt{n}}} \approx$ a standard Normal Z.

$np = 20(\frac{1}{2}) = 0$

$np(1-p) = 5$

Equivalently, $X_1 + X_2 + \cdots + X_{20} \approx$ Normal$(10, 5)$

We know $X_1 + \cdots + X_{20} \sim$ Binomial $(20, \frac{1}{2})$ (see pmf below)



When applying the Central Limit theorem to populations taking integer-values (i.e., discrete populations) the approximation should be corrected by a so-called <u>CONTINUITY</u> <u>CORRECTION</u>

This idea is best seen where the population is a Bernoulli(p).

Suppose $X_1, X_2, \dots, X_{10} \sim \text{Bernoulli}(\frac{1}{2})$ are independent.

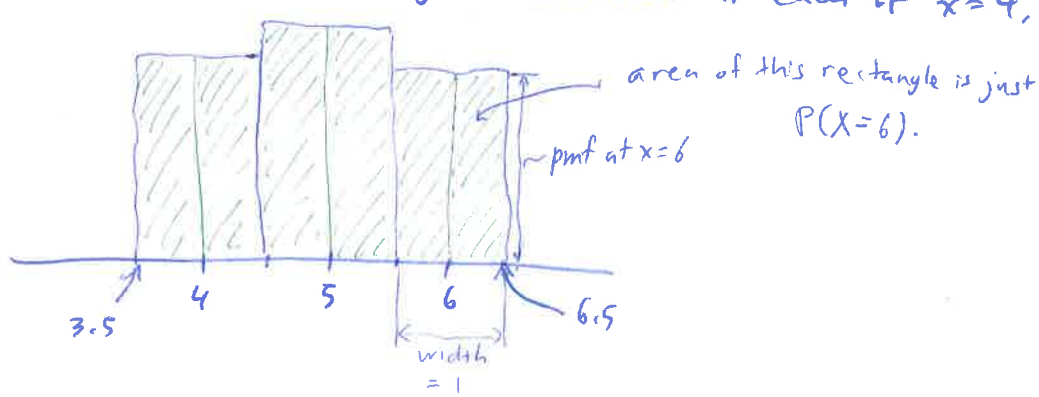
So that $\sum_{i=1}^{10} X_i \sim \text{Binomial}(10, \frac{1}{2})$.

On one hand using the binomial distribution directly:

$$P\left(4 \le \sum_{i=1}^{10} X_i \le 6\right) = \binom{10}{4}\left(\tfrac{1}{2}\right)^{10} + \binom{10}{5}\left(\tfrac{1}{2}\right)^{10} + \binom{10}{6}\left(\tfrac{1}{2}\right)^{10}$$

$$= .65625.$$

i.e., to compute this probability we just add to corresponding probability masses at 4, 5 and 6. We can also think of this as adding the area of the 3 rectangles centered at each of $x=4,5,6$:



To capture the area of all 3 rectangles we should write

$P(4 \le X \le 6)$ as $P(3.5 \le X \le 6.5)$ so that we "pickup" the entire rectangle sentered at 4 and at 6. Otherwise if we performed the standardization without this continuity correction we would get only ½ the area of the rectangles at $x=4$ and $x=6$.

So

$$P(3.5 \le X \le 6.5) = P\left(\frac{3.5-5}{\sqrt{2.5}} \le \frac{X-\mu_X}{\sigma_X} \le \frac{6.5-5}{\sqrt{2.5}}\right)$$

$$\approx P(-.95 \le Z \le .95) = .8289-(1-.8289)$$

$$= .6578$$

Comparing with the exact value of .65625 this is a good approx.

If we had not added ½ and subtracted ½ to capture the entire rectangles centered at $x=4$ and $x=6$ we would have

$$P(4 \le X \le 6) = P\left(\frac{4-5}{\sqrt{2.5}} \le \frac{X-\mu_X}{\sigma_X} \le \frac{6-5}{\sqrt{2.5}}\right)$$

$$\approx P(-.63 \le Z \le .63) = .7357-(1-.7357)$$

$$= .4714.$$

and we see this approximation is not as sharp as with the "Continuity correction".

Moral of the story ...

If we have $X_1, X_2, \ldots, X_n \sim$ iid discrete-integer valued then to estimate $P(a \le \sum_{i=1}^{n} X \le b)$ where a and b are integers make a continuity correction first: ie.

$$P\left(a-\tfrac{1}{2} \le \sum_{i=1}^{n} X_i \le b+\tfrac{1}{2}\right)$$

strict inequality

Also $P\left(a < \sum_{i=1}^{n} X_i < b\right)$ should be continuity-corrected as

$$P\left(a+\tfrac{1}{2} \le \sum_{i=1}^{n} X_i \le b-\tfrac{1}{2}\right)$$ since in this case

$$P\left(a < \sum_{i=1}^{n} X_i < b\right) = P\left(a+1 \le \sum_{i=1}^{n} X_i \le b-1\right)$$

$$= P\left(a+1-\tfrac{1}{2} \le \sum_{i=1}^{n} X_i \le b-1+\tfrac{1}{2}\right)$$

$$= P\left(a+\tfrac{1}{2} \le \sum_{i=1}^{n} X_i \le b-\tfrac{1}{2}\right).$$

$\boxtimes$

Continuing with the above discussion...

If $X_1, X_2, \ldots, X_n \sim$ iid discrete integer-valued r.v.s

then to use the CLT to approximate

$$P(a \le \overline{X}_n \le b)$$

we should continuity correct as follows...

$$P(a \le \overline{X}_n \le b) = P\left(na \le \sum_{i=1}^{n} X_i \le nb\right) = P\left(na-\tfrac{1}{2} \le \sum_{i=1}^{n} X_i \le nb+\tfrac{1}{2}\right)$$

$$= P\left(a-\tfrac{1}{2n} \le \overline{X}_n \le b+\tfrac{1}{2n}\right).$$ Then apply the CLT...

$$P(a \le \overline{X}_n \le b) = P\left(a-\tfrac{1}{2n} \le \overline{X}_n \le b+\tfrac{1}{2n}\right) = P\left(\frac{a-\frac{1}{2n}-\mu}{\sigma/\sqrt{n}} \le \frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}} \le \frac{b+\frac{1}{2n}-\mu}{\sigma/\sqrt{n}}\right)$$

$$\approx P\left(\frac{a-\frac{1}{2n}-\mu}{\sigma/\sqrt{n}} \le Z \le \frac{b+\frac{1}{2n}-\mu}{\sigma/\sqrt{n}}\right) \doteq \Phi\left(\frac{b+\frac{1}{2n}-\mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{a-\frac{1}{2n}-\mu}{\sigma/\sqrt{n}}\right)$$

Example This example I hope demonstrates the power of the CLT and continuity correction.

Early in the course we found that when rolling 3 balanced dice, the probability the sum is 9 $= \frac{25}{216} \approx .1157$.

Let $X_1, X_2, X_3 \sim$ iid discrete uniform on $\{1,2,3,4,5,6\}$.

then we wish to estimate $P(\text{sum}=9) =$

$$P(X_1+X_2+X_3=9) = P(9 \le X_1+X_2+X_3 \le 9)$$

by using the CLT. First we note the population is discrete - integer valued. So we apply a continuity correction:

$P(8.5 \le X_1+X_2+X_3 \le 9.5)$ then use CLT (note how small n is...)

$$P\left(\frac{8.5-3(3.5)}{\sqrt{3(\frac{35}{12})}} \le \frac{X_1+X_2+X_3-n\mu}{\sqrt{n\sigma^2}} \le \frac{9.5-3(3.5)}{\sqrt{3(\frac{35}{12})}}\right)$$

$$\approx P(-.68 \le Z \le -.34) = \Phi(-.34) - \Phi(-.68)$$

$$= (1-\Phi(.34)) - (1-\Phi(.68)) = \Phi(.68) - \Phi(.34)$$

$$= .7517 - .6331 = .1186$$

Compares favorably to the "exact" answer .1157. You should check that without the continuity correction the CLT gives 0 as the approximation.

The Markov inequality.

If $X \geq 0$ is a non-negative random variable and $a > 0$ is any constant, then

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

The statement is true for any r.v. (discrete or continuous) however

The proof of this inequality is surprisingly easy:

Let's assume X is a continuous r.v. with pdf $f(x)$. Then

$$E(X) = \int_0^\infty x f(x)\, dx \geq \int_a^\infty x f(x)\, dx$$

making the interval of integration smaller.

$$\geq \int_a^\infty a f(x)\, dx$$

on the interval $[a, \infty)$ we replace x by the smallest value a.

$$= a \int_a^\infty f(x)\, dx$$

$$= a P(X \geq a).$$

9983

19966

As a consequence of the Markov inequality we obtain the Chebyshev inequality:

for any r.v X with mean μ and variance σ^2, and any $k>0$

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

or equivalently,

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

To see how this follows from the Markov inequality consider the r.v.

$Y = |X-\mu|^2$. This r.v. is clearly nonnegative

and if we take $a = k^2$

$$\underbrace{P(|X-\mu|^2 \geq k^2)}_{= P(|X-\mu| \geq k)} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

$\blacksquare$

Remark about the Chebyshev inequality on previous page

If we take $k = c\sigma$ where $c \geq 1$ (say) and $\sigma = \sqrt{\sigma^2}$,

then the Chebyshev inequality says

$$P(|X-\mu| \geq c\sigma) \leq \frac{\sigma^2}{(c\sigma)^2} = \frac{1}{c^2}.$$

and

$$(*) \quad P(|X-\mu| < c\sigma) \geq 1 - \frac{\sigma^2}{(c\sigma)^2} = 1 - \frac{1}{c^2}.$$

In words, $(*)$ for instance says the probability a r.v. X takes values within c standard deviations of its mean μ is AT Least $1 - \frac{1}{c^2}$.

So that

$$P(-2\sigma < X-\mu < 2\sigma) \geq 1 - \frac{1}{2^2} = \frac{3}{4} = 75\%$$

and $$P(-3\sigma < X-\mu < 3\sigma) \geq 1 - \frac{1}{3^2} = \frac{8}{9} = 88.89\%$$

and $$P(-5\sigma < X-\mu < 5\sigma) \geq 1 - \frac{1}{5^2} = \frac{24}{25} = 96\%$$

at least 96% probability a r.v. X will take values between $\mu - 5\sigma$ and $\mu + 5\sigma$.

Remark

From the previous remark it seems that the 1st two moments of a r.v. put some restrictions on the distribution of the random variable.

Example Let $X_1, X_2, \ldots, X_n$ be a random sample from a population having mean μ and variance σ^2.

I.e. $X_1, X_2, \ldots, X_n \sim$ iid whose prob. distribution with mean μ and var. $= \sigma^2$.

How close is $\bar{X}_n := \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ to μ?

That is, Compute $P(|\bar{X}_n - \mu| < \varepsilon)$ for any small $\varepsilon > 0$.

We'd like to use the Chebyshev inequality here. but we need to find $\mu_{\bar{X}} = E(\bar{X}_n)$ and $\sigma_{\bar{X}}^2 = var(\bar{X}_n)$

$$\mu_{\bar{X}} = E(\bar{X}_n) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}(n\mu) = \mu.$$

$$\sigma_{\bar{X}}^2 = var(\bar{X}_n) = var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \left(\frac{1}{n}\right)^2 var(X_1 + \cdots + X_n)$$

$$= \frac{1}{n^2}\left\{\sum_{i=1}^{n} var(X_i) + \sum_{\substack{i=1 \\ i \neq j}}^{n}\sum_{j=1}^{n} \underbrace{cov(X_i, X_j)}_{\substack{=0 \\ \text{since } i \neq j \\ \text{and } X_i, X_j \\ \text{independent}}}\right\} = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

So, the Chebyshev inequality says

$$P(|\bar{X}_n - \mu_{\bar{X}}| < \varepsilon) \geq 1 - \frac{\sigma_{\bar{X}}^2}{\varepsilon^2}$$

or plugging in what we know about $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}^2$:

$$P(|\bar{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Notice that regardless of how small $\varepsilon > 0$ is, as $n \to \infty$

$\frac{\sigma^2}{n\varepsilon^2} \to 0$. So that

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1.$$

This is called the <u>Weak Law of Lage Numbers</u>.

(or <u>WLLN</u>)

$$P(|\bar{X}_n-\mu|<\varepsilon) = P(-\varepsilon < \bar{X}_n-\mu < \varepsilon)$$ or dividing through by $\frac{\sigma}{\sqrt{n}}$

$$= P\left(-\frac{\varepsilon\sqrt{n}}{\sigma} < \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} < \frac{\varepsilon\sqrt{n}}{\sigma}\right)$$

using the CLT

$$\approx \int_{-\frac{\varepsilon\sqrt{n}}{\sigma}}^{\frac{\varepsilon\sqrt{n}}{\sigma}} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\,du \longrightarrow 1 \text{ as } n\to\infty$$

(since the bounds on the integral are going to $+\infty$ and $-\infty$)

This shows the WLLN can be seen by using the Central limit theorem.