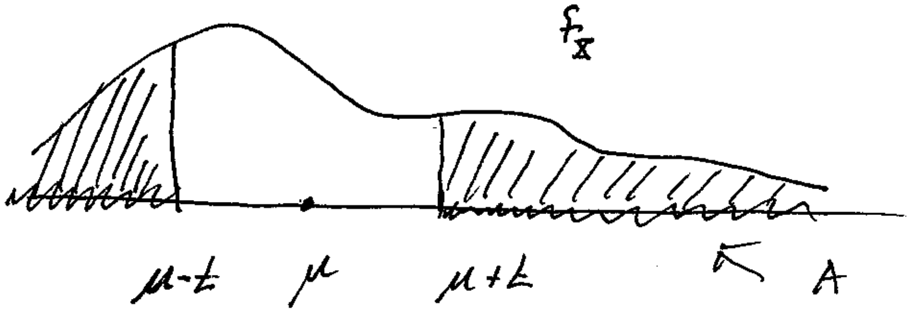
Chebyshev's Inequality

X, μ, σ^2

How likely is X to be near μ?



$$P[|X-\mu| > t] \le \frac{\sigma^2}{t^2}$$

Proof. Let $A = \{x : |x-\mu| > t\}$

$$P[|X-\mu| > t] = \int_A f_X(x)\,dx \le \int_A \boxed{\frac{(x-\mu)^2}{t^2}}\, f_X(x)\,dx$$

≥ 1 on A

$$\boxed{\le \frac{1}{t^2}\int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx}$$

$$= \frac{1}{t^2}\, E(X-\mu)^2$$

$$= \frac{1}{t^2} \cdot \sigma^2$$

Discrete case is similar.

Law of Large Numbers

$X_1, X_2, \ldots, X_i, \ldots$

independent, $E(X_i) = \mu$ $\operatorname{var} X_i = \sigma^2$

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Let $\epsilon > 0$. Then

$$P\left[\,|\bar{X}_n - \mu| > \epsilon\,\right] \to 0 \quad \text{as } n \to \infty$$

Proof: $E(\bar{X}_n) = \frac{1}{n} \cdot n\mu = \mu$

$$\operatorname{var} \bar{X}_n = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

By Chebyshev's inequality,

$$P\left[|\bar{X}_n - \mu| > \epsilon\right] \leq \frac{\sigma^2}{n \cdot \epsilon^2} \to 0$$

as $n \to \infty$.

example: Toss a fair coin;

$$P(\text{"heads"}) = \frac{1}{2}$$

Long run approach:

$$\boxed{0,1} \curvearrowright X_i$$

$$X_1, X_2, \ldots, X_n$$

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \text{"proportion of heads in } n \text{ tosses"}$$

$$E(\bar{X}_n) = \frac{1}{n} \cdot n \cdot \frac{1}{2} = \frac{1}{2}$$

law of large numbers $\Rightarrow$ given $\epsilon > 0$,

$$P\left[\,\left|\bar{X}_n - \frac{1}{2}\right| > \epsilon\,\right] \longrightarrow 0$$

as $n \to \infty$

$$P\left[\,\left|\bar{X}_n - \frac{1}{2}\right| \le \epsilon\,\right] \longrightarrow 1$$

as $n \to \infty$.

"Life is like *drawing from* a box of chocolates."
Sometimes you have a pretty good idea of what you are going to get.

I. NORMAL APPROXIMATION THEOREM

[...21, 3, 2, -435, 61, 1943, $\sqrt{2}$, e , π, ...]

Draw n times with replacement.

Suppose you draw n times, with replacement from a box of numbers (not all zero). Let X denote the sum (or average) of the numbers drawn. For n large, the probability density function for $\frac{X-\mu_X}{\sigma_X}$ is approximately N(0,1) .

II. CENTRAL LIMIT THEOREM

Let $X_1, X_2, X_3, \ldots$ be a sequence of independent and identically distributed random variables each having mean μ and variance σ^2. Then the distribution of

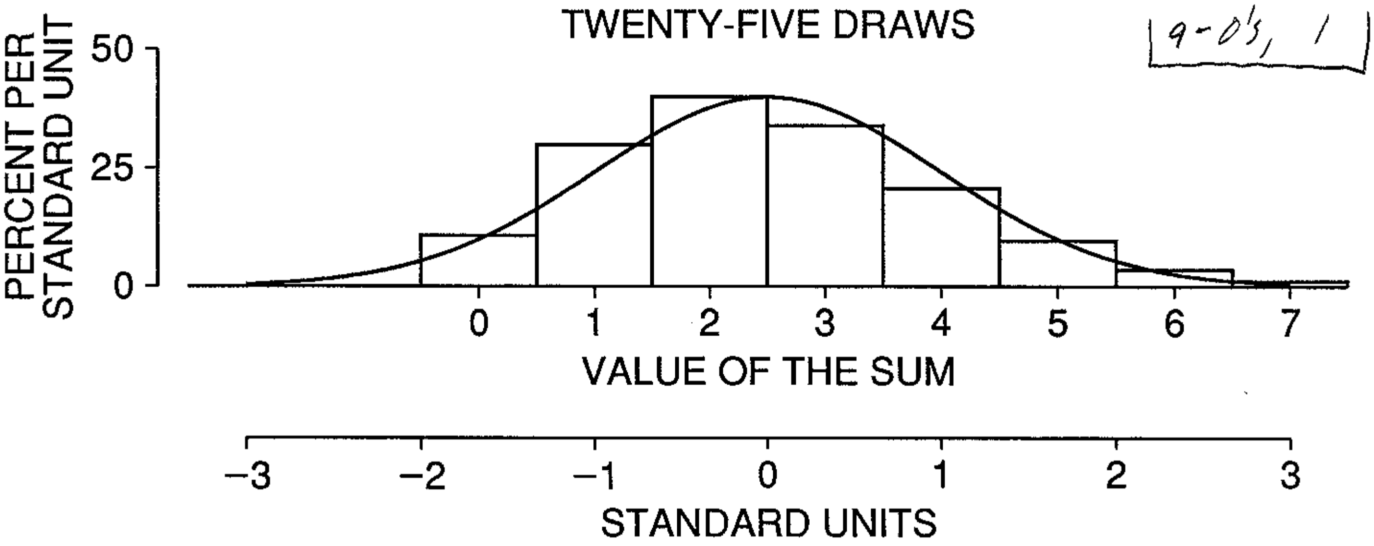
$$\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$.

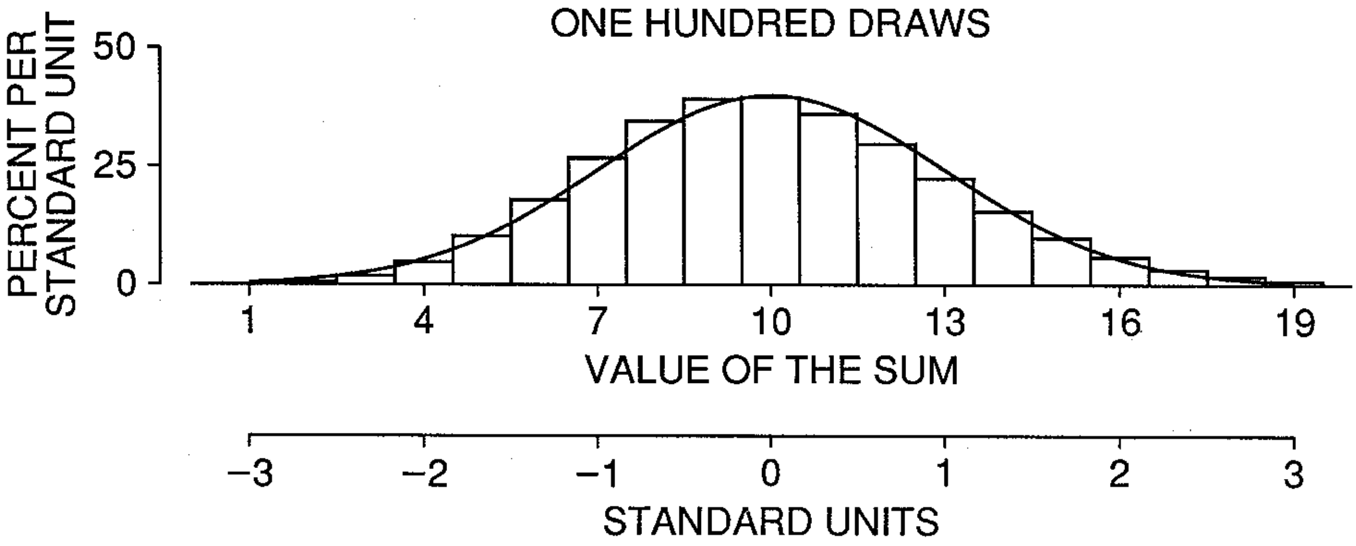
That is, as $n \to \infty$

$$P\left\{ \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \leq a \right\} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx$$

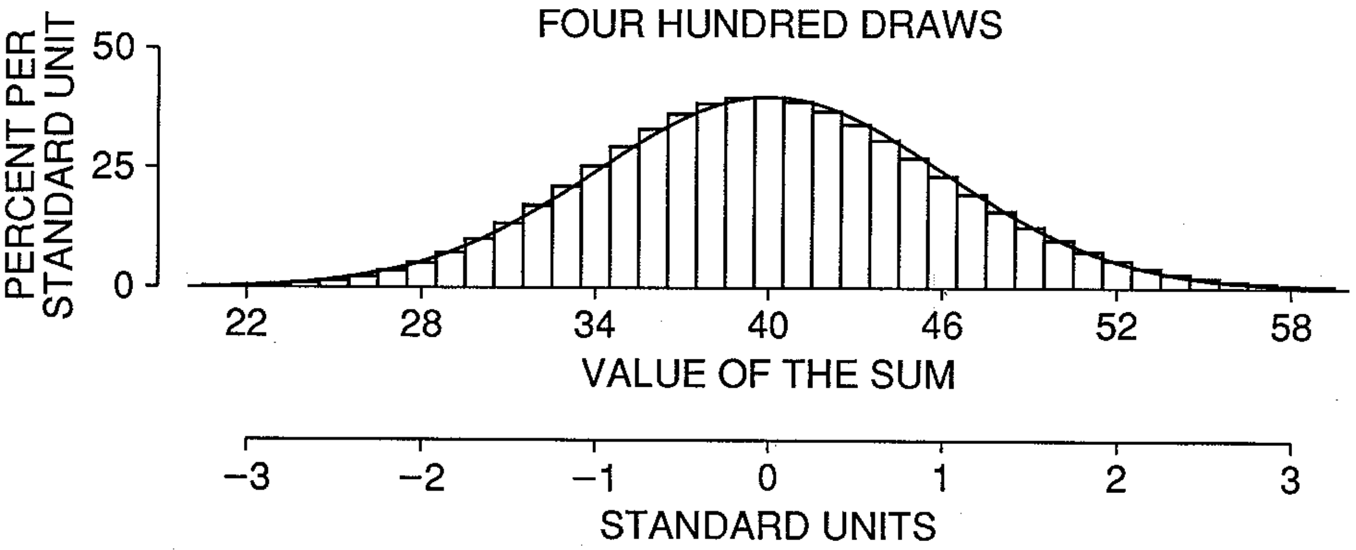
9-0's, 1



TWENTY-FIVE DRAWS



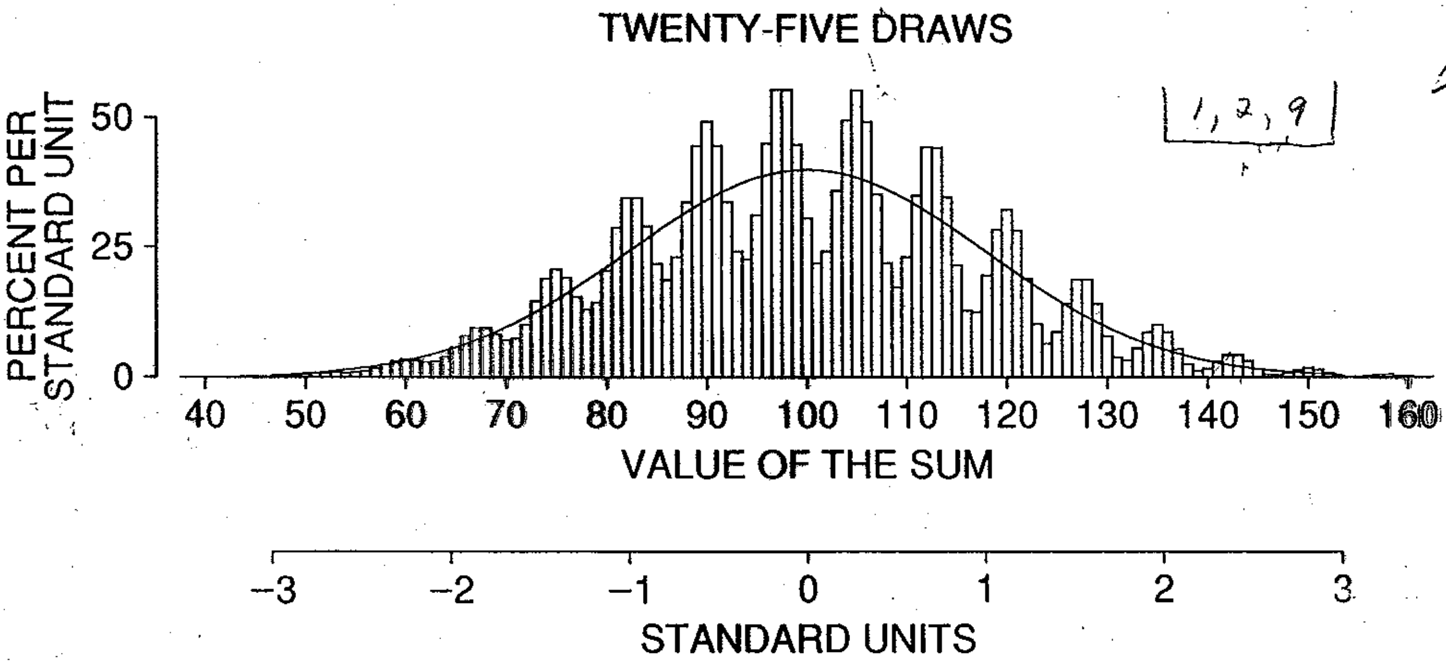
ONE HUNDRED DRAWS

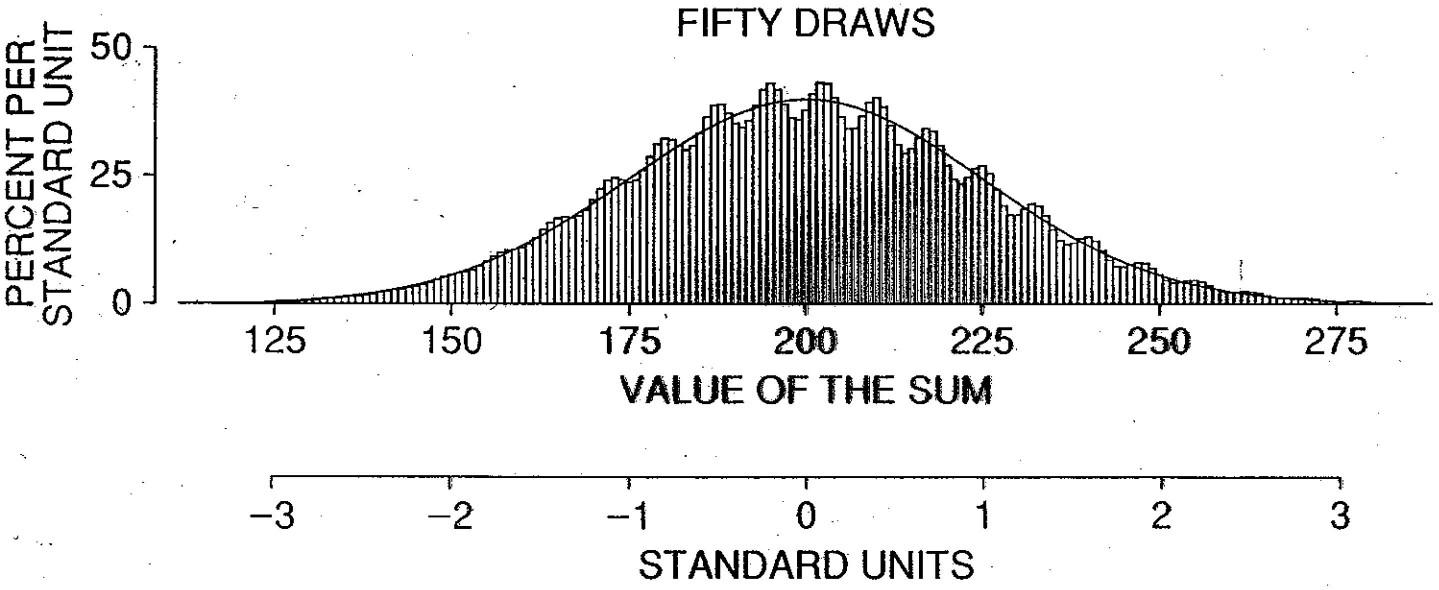


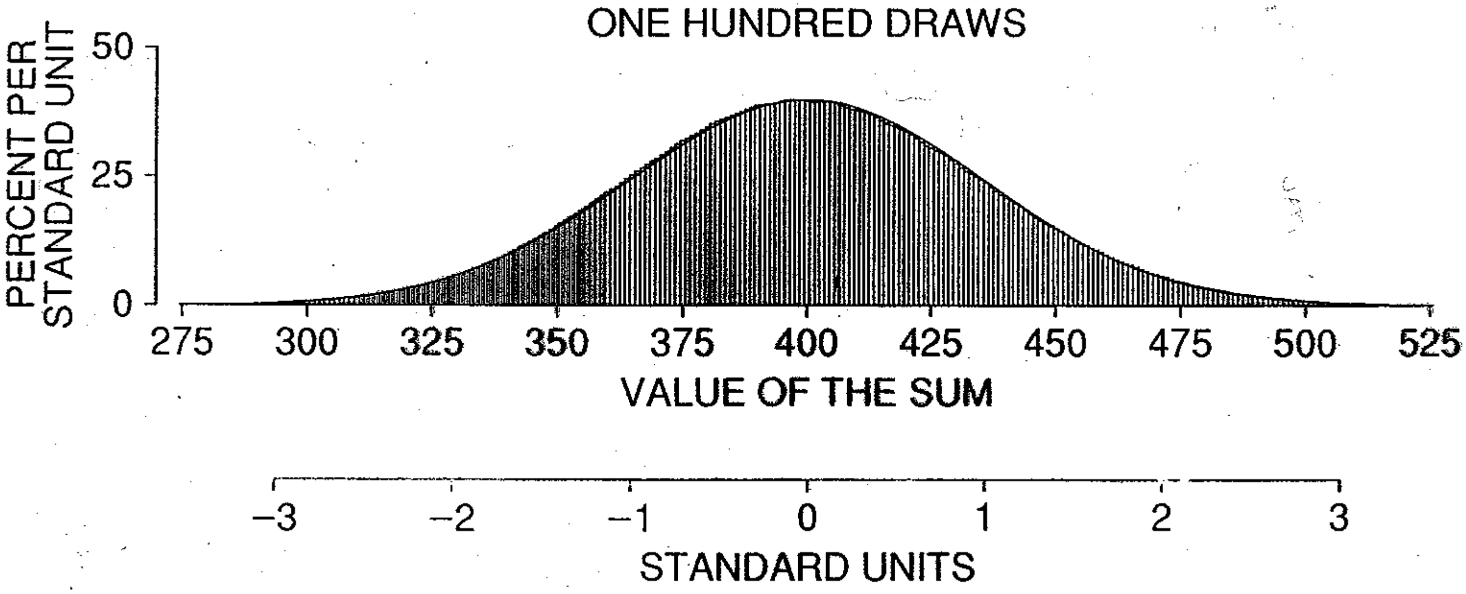
FOUR HUNDRED DRAWS

34-7

1, 2, 9


TWENTY-FIVE DRAWS
PERCENT PER STANDARD UNIT
50
25
0
40
50
60
70
80
90
100
110
120
130
140
150
160
VALUE OF THE SUM
–3
–2
–1
0
1
2
3
STANDARD UNITS


FIFTY DRAWS
PERCENT PER STANDARD UNIT
50
25
0
125
150
175
200
225
250
275
VALUE OF THE SUM
–3
–2
–1
0
1
2
3
STANDARD UNITS


ONE HUNDRED DRAWS
PERCENT PER STANDARD UNIT
50
25
0
275
300
325
350
375
400
425
450
475
500
525
VALUE OF THE SUM
–3
–2
–1
0
1
2
3
STANDARD UNITS

$X \sim$ Binomial, n, p

$$M_X(t) = (1-p+pe^t)^n$$

$Y \sim$ Poisson, λ

$$M_Y(t) = e^{\lambda(e^t-1)}$$

Suppose $n \to \infty$, $p \to 0$, $np = \lambda$, $p = \frac{\lambda}{n}$

$$\lim_{n\to\infty} M_X(t) = \lim_{n\to\infty} \left(1-\frac{\lambda}{n}+\frac{\lambda}{n}e^t\right)^n$$

Let $w(n) = \left(1-\frac{\lambda}{n}+\frac{\lambda}{n}e^t\right)^n$. Now

$\lim_{n\to\infty} w(n) = ?$ Consider $\ln[w(n)]$

$$\lim_{n\to\infty} \ln[w(n)] = \lim_{n\to\infty} n \cdot \ln\left(1-\frac{\lambda}{n}+\frac{\lambda}{n}e^t\right)$$

$$= \lim_{n\to\infty} \frac{\ln\left(1-\frac{\lambda}{n}+\frac{\lambda}{n}e^t\right)}{\frac{1}{n}} = \lim_{n\to\infty} \frac{1}{1-\frac{\lambda}{n}+\frac{\lambda}{n}e^t}(-\lambda+\lambda e^t)$$

$$= \cancel{\lim_{n\to\infty}} = -\lambda+\lambda e^t.$$ So

$$\lim_{n\to\infty} w(n) = e^{-\lambda+\lambda e^t} = [illegible]$$

$$= e^{\lambda(e^t-1)}$$

Math 5710

1. Let X be equal to the sum of 400 draws (with replacement) from the box [0, 0, 0, 1]. Find the expected value of X and the standard deviation of X.

2. A certain town has 25,000 families. The average number of children per family is 2.6 with an SD of 0.8. The distribution is not normal, however, since 25% of the families have no children at all. If we draw a random sample of 400 families, what are the chances that between 23% and 27% of the sample families will have no children?

3. Toss a quarter 10 times. Suppose you observed only two heads. Is the coin fair? This result could be just due to chance variation. How likely is it to observe two or fewer heads in ten tosses of a fair coin? Let $X =$ the number of heads in ten tosses of a fair coin. Find $p = P(X \leq 2)$. What do you conclude?

3. I tossed my Utah quarter 400 times and got a total of 225 "golden spikes". Assuming that Utah quarters are fair, use the Central Limit Theorem to approximate the probability that I would get 225 or more "golden spikes" in 400 tosses. Is my Utah quarter fair?

4. Three hundred draws will be made at random with replacement from the box [0 , 0 , 0 , 1, 1] . For each i, let X_i be the number drawn on the i-th draw and let

$$\bar{X} = \frac{1}{300} \sum_{i=1}^{300} X_i \, .$$

Use the Central Limit Theorem to approximate $P(\bar{X} > 0.428)$.

Math 5710
Example

An astronomer is interested in measuring, in light years, the distance from his observatory to a distant star. Although the astronomer has a measuring technique, she knows that, because of changing atmospheric conditions and normal error, each time a measurement is made it will not yield the exact distance but merely an estimate. As a result the astronomer plans to make a series of measurements and then use the average value of these measurements as her estimated value of the actual distance. If the astronomer believes that the values of the measurements are independent and identically distributed random variables having a common mean *d* (the actual distance) and a common variance of 4 (light years), how many measurements need she make to be reasonably sure that her estimated distance is accurate to within ½ light year?

X_i = i-th measurement $\quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Assume $E(X_i) = d$ light years.

var $X_i = \sigma^2 = 4$ for all i.

$E(\bar{X}_n) = d$, var $\bar{X}_n = \frac{1}{n^2}\cdot n \cdot 4 = \frac{4}{n}$

$\sigma_{\bar{X}_n} = \frac{2}{\sqrt{n}}$.

Want $P\left[d-\frac{1}{2} < \bar{X}_n < d+\frac{1}{2}\right] = \boxed{.95}$

$$P\left[-\frac{1}{2} < \bar{X}_n - d < \frac{1}{2}\right] = .95$$

$$P\left[-\frac{1}{2}\cdot\frac{\sqrt{n}}{2} < \frac{\bar{X}_n - d}{2/\sqrt{n}} < \frac{1}{2}\cdot\frac{\sqrt{n}}{2}\right] = .95$$

Central Limit Theorem $\Rightarrow$

$\frac{\sqrt{n}}{4} = 2$, $\sqrt{n} = 8$, $\boxed{n = 64}$

4. Consider the following box (population) of 0's and 1's: [? 0's , ? 1's]. The number of 0's in the box is unknown and the number of 1's is also unknown. Let $\mu =$ Box AV $=$ proportion of 1's. Let $\sigma =$ Box SD.

What is μ? Can we estimate the value of μ?

Let $X_1, X_2, \ldots, X_n$ be a random sample from the box (with replacement). Then

$\bar{X}_n =$ $W_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ is a reasonable estimator for μ. Note that $E(W_n) = \mu$.

Sometimes $W_n < \mu$ and sometimes $W_n > \mu$ but on average $W_n = \mu$. However, W_n is useless as an estimator of μ unless we can determine how accurate it is.

We know that W_n is approximately a normal distribution for n large.

So, $P\left(-2 < \dfrac{W_n - \mu}{\sigma/\sqrt{n}} < 2\right) \approx .95$

Then, $P\left(\dfrac{-2\sigma}{\sqrt{n}} < W_n - \mu < \dfrac{2\sigma}{\sqrt{n}}\right) \approx .95$

and so, $P\left(W_n - \dfrac{2\sigma}{\sqrt{n}} < \mu < W_n + \dfrac{2\sigma}{\sqrt{n}}\right) \approx .95.$

Since $\sigma \le \dfrac{1}{2}$, we have

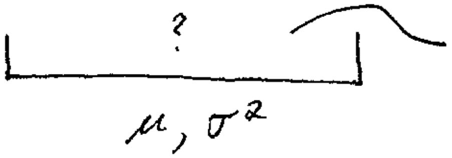
$P\left(W_n - \dfrac{1}{\sqrt{n}} < \mu < W_n + \dfrac{1}{\sqrt{n}}\right) \approx .95$ or more

Suppose $n = 400$ and $W_n = .65$. Then,

$$\left(.65 - \frac{1}{20}\ ,\ .65 + \frac{1}{20}\right) = (.6\ ,\ .7)$$

is a 95% confidence interval for μ. We now have obtained a measure of how accurate our estimator W_n is of μ.

Statistical Sampling



$X_1, X_2, \ldots, X_n$

independent

"random sample"

$\forall i, \quad E(X_i) = \mu \quad , \quad \text{var}\, X_i = \sigma^2$

① How do we estimate μ ?

"statistic" $\quad \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n} \qquad E(\bar{X}) = \mu$

"unbiased"

Note that

$\text{var}\, \bar{X} = E(\bar{X} - \mu)^2 = \frac{1}{n^2} \cdot n \sigma^2$

$= \frac{\sigma^3}{n}$

② How do we estimate σ^2 ?

a) Suppose we know μ. Let

$$W = \frac{1}{n} \sum_{i=1}^{n} (X_i - \underset{\uparrow}{\mu})^2 \; .$$

$$E(W) = \frac{1}{n} \cdot n\, \sigma^2 = \sigma^2$$

"unbiased"

b) What if we do not know μ? Let

$$V = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 \qquad \boxed{\frac{\sum (X_i-\bar{X})^2}{n-1}}$$

$$E(V) = \frac{1}{n}E\sum_{i=1}^{n}\left[(X_i-\mu)-(\bar{X}-\mu)\right]^2$$

$$= \frac{1}{n}E\left\{\sum_{i=1}^{n}\left[(X_i-\mu)^2-2(X_i-\mu)(\bar{X}-\mu)+(\bar{X}-\mu)^2\right]\right.$$

$$= \frac{1}{n}E\left\{\sum_{i=1}^{n}(X_i-\mu)^2-2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu)+\sum_{i=1}^{n}(\bar{X}-\mu)^2\right.$$

$$= \frac{1}{n}E\left\{\sum_{i=1}^{n}(X_i-\mu)^2-2(\bar{X}-\mu)\cdot\left[\sum_{i=1}^{n}X_i-n\mu\right]+n(\bar{X}-\mu)^2\right\}$$

$$= \frac{1}{n}E\left\{\sum_{i=1}^{n}(X_i-\mu)^2-2(\bar{X}-\mu)\cdot n\left[\frac{\sum X_i}{n}-\mu\right]+n(\bar{X}-\mu)^2\right\}$$

$$= \frac{1}{n}E\left\{\sum_{i=1}^{n}(X_i-\mu)^2-2(\bar{X}-\mu)\cdot n(\bar{X}-\mu)+n(\bar{X}-\mu)^2\right.$$

$$= \frac{1}{n}E\left\{\sum_{i=1}^{n}(x_i-\mu)^2-n(\bar{X}-\mu)^2\right\}$$

$$= \frac{1}{n}\left[n\sigma^2-\sigma^2\right] = \boxed{\frac{n-1}{n}}\sigma^2 \qquad \text{biased!!}$$