Announcements

Lab 1 due February 5, 2010
Lab Clues

Before lab:
1. Read over the lab.
2. Highlight all the verbs in the procedure section.
3. Do a preliminary write up of objective, and idea.

During 1st lab period for each experiment:
1. Take all the data if possible.
2. Make notes of all errors you can think of as you go.
3. If there are any notes by the apparatus, write them down!
Lab Clues

Between 1\textsuperscript{st} and 2\textsuperscript{nd} lab periods:
1. Do as much analysis as possible:
   a) to find out if you need to take more data
   b) to figure out what questions you will need to ask in lab
2. Get as far through the lab report as you can.

After 2\textsuperscript{nd} lab period:
1. You will write up final report. Since you will have questions, be sure you get them answered before now!
2. Labs are due Feb. 5.
• **Uncertainty in Scientific Measurements**
• **The Gaussian Distribution**
  • What it looks like
  • Where and why it shows up
  • Mean, sigma, and all that
• **Random and Systematic Error**
UNCERTAINTY IN SCIENTIFIC MEASUREMENTS

Measure the height of a doorway:
- Estimate: 210 cm
- Tape measure: 211.3 cm
- Laser measure: 211.3065 cm
- What is the exact height? What does that even mean?
- No physical quantity can be measured with absolute certainty!
UNCERTAINTY IN SCIENTIFIC MEASUREMENTS

- Is King’s crown made of gold ($\rho_g=15.5 \text{ g/cm}^3$) or alloy ($\rho_a=13.8 \text{ g/cm}^3$)?

- **George:** $\rho=15$, probably between 13.5-16.5 g/cm$^3$
- **Martha:** $\rho=13.9$, probably between 13.7-14.1 g/cm$^3$

**Both may be reasonable measurements.**

- **George** is less precise, can’t conclude. If no uncertainty → gold!
- **Martha** must justify her range for us to believe her conclusion of alloy.
Quiz

Q: From these measurements which conclusion would you draw about the crown?

A. It is made of gold.
B. It is made of an alloy.
C. You can’t tell because the measurements are inconsistent.
D. The measurements are consistent, but you still can’t tell.
UNCERTAINTY IN SCIENTIFIC MEASUREMENTS

- **Measure circumference** of 150 cows.

- The values appear to be centered on the **average value** (mean, \( \mu \)).

- The values appear to occur more frequently near the average and less frequently further away. (Symmetric distribution.)

- **Mean** and width (standard deviation, \( \sigma \)) describe distribution.
GAUSSIAN (NORMAL) DISTRIBUTION

- **Synonyms**: Normal Distribution, Bell Curve

- Most basic form is “unit” Gaussian: centered at zero, unit integral, unit $\sigma$:

  $$F(x) = \sqrt{\frac{1}{2\pi}} \exp \left( -\frac{x^2}{2} \right)$$

- This is the **probability density** for a continuous, normally distributed random variable with mean zero and standard deviation of 1.

  As with any probability density function, 

  $$\int_{-\infty}^{+\infty} dx F(x) = 1.$$
GAUSSIAN DISTRIBUTION

- What can you do to the unit gaussian, and still keep it a gaussian?
  - Change its mean from zero $\rightarrow \mu$
  - Change its width from 1 $\rightarrow \sigma$
  - Change its integral — but then it’s not a normalized probability distribution anymore. So don’t allow this!

$$F(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[\frac{-(x - \mu)^2}{2\sigma^2}\right]$$

As with any probability density function, this still integrates to 1.
WHERE IT SHOWS UP

• If you don’t have a clue what the probability distribution of a random quantity is (say, the circumferences of cows), it’s highly likely to be approximately gaussian!

• This is due to the Central Limit Theorem: a sum of a large enough number of random numbers has a gaussian distribution, no matter what the initial distribution shapes might have been.

• Aside: Distribution of counts of a process with a uniform rate in a finite amount of time is Poisson-distributed (see a later lecture) but is approximately gaussian in the high-number limit. This is another example of the Central Limit Theorem.

• The mean is the best measurement of the true value.
MEAN, SIGMA, AND ALL THAT

- Say we measured 150 cows, and have the histogram of results.

- Calculate the mean circumference $<C>$:

$$<C> = \frac{1}{N} \sum_{i} c_i = 4.39 \text{ m}$$

- Standard deviation is calculated by:

$$\sigma_c = \sqrt{\frac{\sum_{i=1}^{N} (c_i - <C>)^2}{N - 1}} = 0.70 \text{ m}$$

- Read up in Taylor on definitions and uses of variance, standard deviation.
FLIPPING COINS

\[ h + t = 1 \]

two flips: hh, ht, tt
1, 2, 1
1/4, 2/4, 1/4

(long table)

three flips: hhh, hht, thh, ttt
1, 3, 3, 1
1/8, 3/8, 3/8, 1/8

(long table)

\[
(h + t)^2(h + t) = h^2 + 2ht + t^2
\]

\[
(h + t)^3(h + t)^2 = h^3 + 3h^2t + 3ht^2 + t^3
\]
BINOMIAL EXPANSION

\[(a + b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \binom{n}{3}a^{n-3}b^3 + \ldots + \binom{n}{n}b^n\]

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}
\]

\[
P(N, n) = \frac{N!}{(N-n)!n!} h^n (1-h)^{N-n}
\]

\[
P(N, n) = \frac{N!}{(N-n)!n!} 2^{-N}
\]
OUTCOME PROBABILITY IS GAUSSIAN (NORMAL) DIST.
1 the gaussian approximation to the binomial

we start with the probability of ending up $j$ steps from the origin when taking a total of $N$ steps, given by

$$P_j = \frac{N!}{2^N \left( \frac{N+j}{2} \right)! \left( \frac{N-j}{2} \right)!}$$  \hspace{1cm} (1)

taking the logarithm of both sides, we have

$$\ln P_j = \ln N! - N \ln 2 - \ln \left( \frac{N+j}{2} \right)! - \ln \left( \frac{N-j}{2} \right)!$$  \hspace{1cm} (2)

now we apply stirling’s approximation, which reads

$$\ln N! \approx N \ln N - N + \frac{1}{2} \ln 2\pi N$$  \hspace{1cm} (3)

for large $N$, this gives

$$\ln P_j \approx \left[ N \ln N - N + \frac{1}{2} \ln 2\pi N \right] - N \ln 2 -$$

$$\left[ \left( \frac{N+j}{2} \right) \ln \left( \frac{N+j}{2} \right) - \left( \frac{N+j}{2} \right) + \frac{1}{2} \ln 2\pi \left( \frac{N+j}{2} \right) \right]$$

$$- \left[ \left( \frac{N-j}{2} \right) \ln \left( \frac{N-j}{2} \right) - \left( \frac{N-j}{2} \right) + \frac{1}{2} \ln 2\pi \left( \frac{N-j}{2} \right) \right]$$
the second term in each of the square brackets cancel each other. regrouping the first term in each of the square brackets together,

$$\ln P_j \approx \left[ N \ln N - \left( \frac{N + j}{2} \right) \ln \left( \frac{N + j}{2} \right) - \left( \frac{N - j}{2} \right) \ln \left( \frac{N - j}{2} \right) \right]$$

$$+ \frac{1}{2} \ln 2\pi N - \frac{1}{2} \ln 2\pi \left( \frac{N + j}{2} \right) - \frac{1}{2} \ln 2\pi \left( \frac{N - j}{2} \right) - N \ln 2$$

looking at only the terms in square brackets and rearranging a bit,

$$\left[ \quad \right] = N \ln N - \frac{N}{2} \ln \left( \frac{N + j}{2} \right) \left( \frac{N - j}{2} \right) - \frac{j}{2} \ln \left( \frac{N + j}{2} \right) + \frac{j}{2} \ln \left( \frac{N - j}{2} \right)$$

$$= N \ln N - \frac{N}{2} \ln \left( \frac{N^2}{4} \left( 1 - \frac{j^2}{N^2} \right) \right) - \frac{j}{2} \ln \left( \frac{N}{2} (1 + \frac{j}{N}) \right) + \frac{j}{2} \ln \left( \frac{N}{2} (1 - \frac{j}{N}) \right)$$

$$= N \ln N - \frac{N}{2} \ln \left( \frac{N^2}{4} \right) - \frac{N}{2} \ln (1 - \frac{j^2}{N^2}) - \frac{j}{2} \ln (1 + \frac{j}{N}) + \frac{j}{2} \ln (1 - \frac{j}{N})$$

now we use the taylor expansion \( \ln(1 \pm x) \approx \pm x \) for \( x \ll 1 \) and work to “second order in \( \frac{j}{N} \)":

$$\left[ \quad \right] \approx N \ln 2 + \frac{N}{2} \frac{j^2}{N^2} - \frac{j}{2} \frac{j}{N} - \frac{j}{2} \frac{j}{N}$$

$$= N \ln 2 - \frac{j^2}{2N}$$

simplifying Eqn. (4) a bit and plugging the above in gives

$$\ln P_j \approx \left[ N \ln 2 - \frac{j^2}{2N} \right] - \frac{1}{2} \ln \left[ \frac{2\pi}{N} \left( \frac{N + j}{2} \right) \left( \frac{N - j}{2} \right) \right] - N \ln 2$$
AN EXAMPLE OF THE CENTRAL LIMIT THEOREM

\begin{align*}
\ln P_j & \approx -\frac{j^2}{2N} - \frac{1}{2} \ln \left[ \frac{\pi}{2N} N^2 (1 - \frac{j^2}{N^2}) \right] \\
& \approx -\frac{j^2}{2N} - \frac{1}{2} \ln \frac{\pi N}{2}
\end{align*}

we exponentiate to get $P_j$ back:

$$P_j \approx \sqrt{\frac{2}{\pi N}} e^{-\frac{j^2}{2N}}$$

(7)

lesson: the large-$N$ limit of a “fair” binomial distribution is a gaussian distribution.
MEAN, SIGMA, AND ALL THAT

- So, we can now say that the mean circumference is $4.39 \text{ m}$ and the standard deviation is $0.70 \text{ m}$. What do we know?

- 68% of cows have circumference between $(4.39 - 0.70)$ and $(4.39 + 0.70)$ m. This is because the integral of the gaussian from $\mu - \sigma$ to $\mu + \sigma$ is 0.68.

\[ \int_{\mu - \sigma}^{\mu + \sigma} dx F_{\mu, \sigma}(x) = 0.68 \]

- How well do we know the mean circumference? Need std. dev. on the mean:

\[ \sigma_{\langle c \rangle} = \frac{\sigma_c}{\sqrt{N}} = 0.06 \]
MEAN, SIGMA, AND ALL THAT

- So $\langle c \rangle = (4.39 \pm 0.06) \text{ m}$, where “±” means 1 sigma, or 68% probability that the true mean is within that interval (68% “confidence level”).

- However, here we had 150 measurements. If we only have 5-10 measurements, the std. dev., $\sigma_C$ is an appropriate approximation to the std. dev. on the mean, $\sigma_{\langle c \rangle}$.

- Std. dev. on the mean is a theoretical concept. In the real world, you will almost always see people use std. dev. And almost never see people use std. dev. on the mean. In this class, always use std. dev.
Can use the same mathematics to describe the results of repeated measurements of the same quantity, where there is random error/resolution in the instrument.

The distribution will be centered on a mean (assume for now that this is the correct value)

The distribution will have a standard deviation

Can fit this to a gaussian or just calculate mean, sigma directly

Uncertainty on the mean is now \( \sigma_{\mu} = \frac{\sigma}{\sqrt{N}} \)

Note: More measurements → smaller error on the mean! Also means better determination of error. For a small number of measurements this does not make sense.