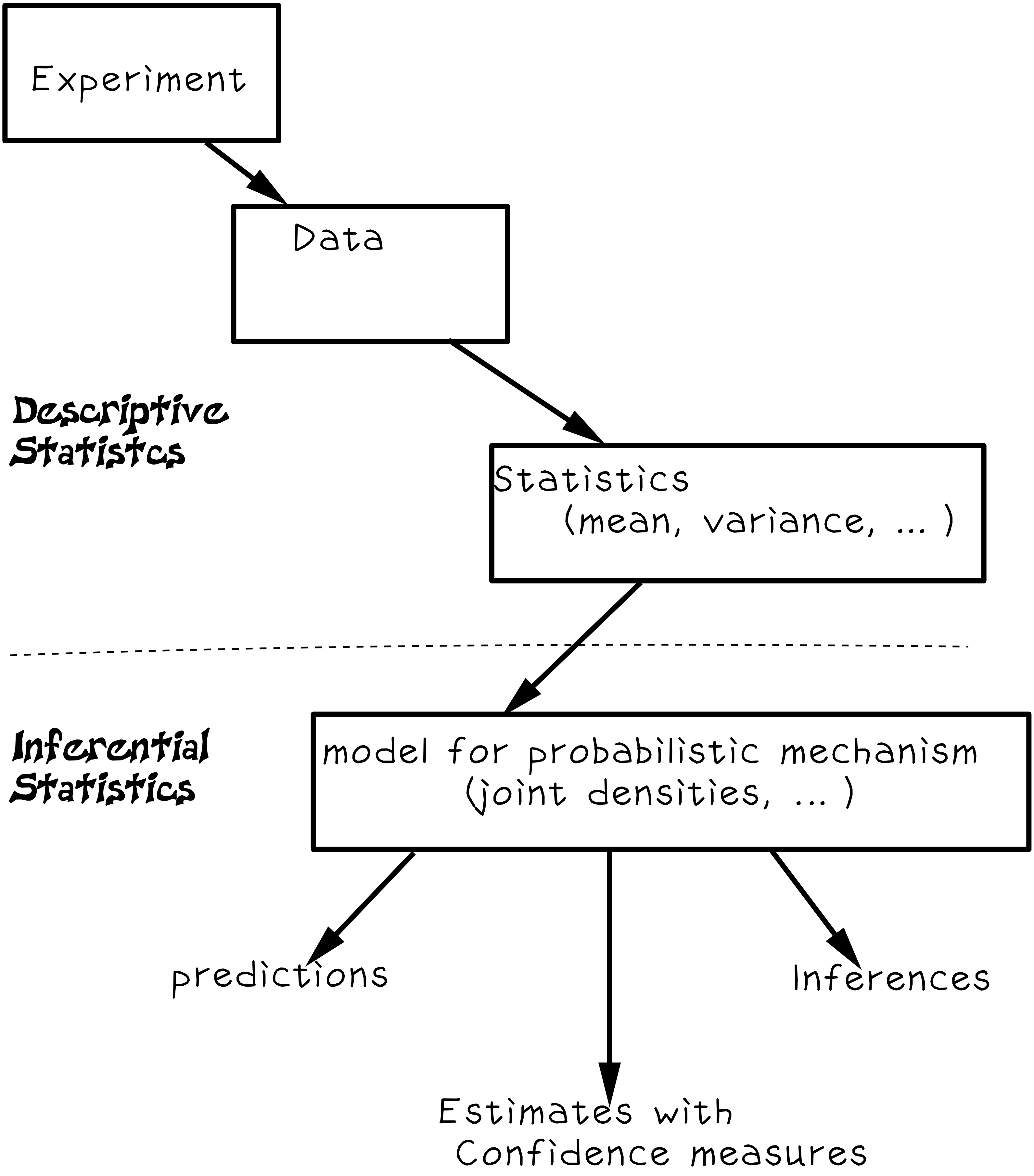
Statistics

Descriptive and Inferential Statistics

Estimating σ and μ

Confidence Intervals

Simulating random arrivals


Experiment
Data
Descriptive Statistcs
Statistics (mean, variance, ...)
Inferential Statistics
model for probabilistic mechanism (joint densities, ...)
predictions
Inferences
Estimates with Confidence measures

Let x be a random variable of known probability distribution.

MEAN of x: $\mu = E(x) = \int xp(x)\ dx$

VARIANCE of x:

$$\sigma^2 = E[(x - \mu)^2] =$$
$$E[x^2] - 2\,E[x\mu] + E[\mu^2] =$$
$$E[x^2] - \mu^2$$

σ is the standard deviation

Estimating from independent observations x1,x2, ... xn.

Estimate of mean $\bar{x}$ 1/n(x1 + x2 + ... + xn) is called the sample mean.

Estimate of variance

$$s^2 = 1/(n-1) \sum (xi - \bar{x})^2 \quad \text{(Why n-1??)}$$

Theorem: x is an unbiased estimate of μ and s^2 is an unbiased estimate of σ^2.

Importance of the Gaussian Distribution

Sum of independent observations converge to this limit -- Central Limit Theorem...

Useful for measuring data as well as the noise in data.

Linear combinations of Gaussian's are Gaussian

Has maximum entropy (ie is most random)

The sample mean and sample variance of N independent, identically distributed samples are independent. The sample mean is also normal w/ the same mean as the parent distribution and 1/Nth the variance.

Distributions derived from Gaussian

Certain statistics derived from independent samples of a Gaussian have special known distributions which can be computed fairly easily. (They used to be published in big fat tables.)

Two most important

normalized sample variance (sum of the square deviations from sample mean, divided by the true variance) of N independent samples from a Gaussian is **chi squared** distributed with N-1 degrees of freedom

deviation of sample mean from true mean, normalized appropriately, is **student t** distributed with N-1 degrees of freedom.

Confidence intervals, interpretation of results

How to measure the accuracy of your estimates? (the world of error bars)

Assume you are measuring independent samples from a Gaussian, the distributions above enable you to estimate confidence intervals.

For example, with sample mean....

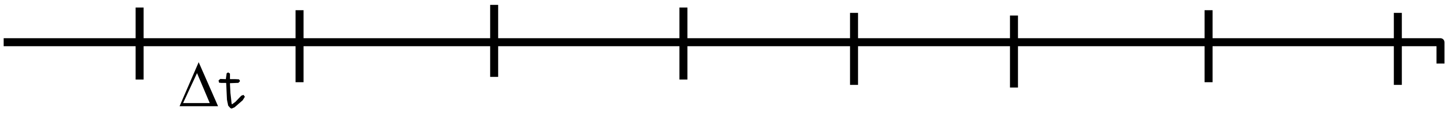
Normalized deviation from true mean has a known distribution (**student t**).

Look up 2 values, say L and R, with the property that the probability that your observation is within the interval [L,R] is 99%.

Plot this interval around your data point, giving the viewer an estimate of how accurate the observation is.

For sample variance, do the same but replace **student t** by **chi squared** distribution.

Simulating random arrivals (Back to car wash)



Assume probability of arrival in any interval
a) proportional to interval length $(\mu\Delta t)$
b) independent of arrivals in other intervals
c) small so can ignore prob. of >1 arrivals in Δt

prob(no arrivals in 1st n intervals) = $(1 - \mu\Delta t)^n$

prob(no arrival to time t) = $(1 - \mu\Delta t)^n$ =

$$\left((1 - \mu\Delta t)^{1/\mu\Delta t}\right)^{\mu t}$$

Now $n = t/\Delta t$ and $\lim_{\Delta t \to 0} (1-x)^{1/x} = e^{-1}$

So, prob(no arrivals until t) = $e^{-\mu t}$

prob(1st arrival before t) = $1 - e^{-\mu t}$

prob. density (time t to first event) =

$$d/dt(1 - e^{-\mu t}) = \mu e^{-\mu t}$$

Random variable with this distribution is said to be Poisson

For discussion

course mailing list

how much cooperation should be allowed?

thursday assignment due -- Dannie to lead discussion

any questions?

precept on weds will be fun.