Course Materials

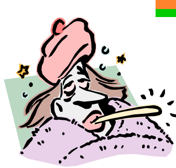
- TEXT:
  - Applied Probability and Statistics for Engineers by D.C.Montgomery and G.C.Runger (John Wiley & Sons, 1999)
- http://www.lehigh.edu/~eup2/teaching/ie121
  - Announcements
  - Lecture notes
  - Homeworks and solutions, and more...

Course Outline

- Probability Review
- Parameter Estimation
- Hypothesis Testing, Statistical Inference
- Regression and Correlation
- Analysis of Variance
- Non-Parametric Statistics
- Statistical Quality Control

Probability Review

- FOUR Laws of Probability:
  - $0 \leq P(A) \leq 1$
  - If A,B are m.e., then $P(A \cup B) = P(A) + P(B)$
  - If A,B are indept, then $P(A \cap B) = P(A).P(B)$
  - $P(A|B) = P(A \cap B) / P(B)$. By extension, Bayes' Law: $P(A|B) = P(A).P(B|A) / P(B)$



Probability Example 1

- A medical test for malaria is subject to both false-positive and false-negative errors. Given that a person has malaria, the probability the test will fail to reveal it is 0.06. And, given that a person does not have malaria, the chance is 0.09 that the test will suggest the opposite.

Probability Example 1...

- Given that Guido has malaria, what is the chance that his test result will reflect it?
- From earlier information, a physician concludes that Guido has a 70% chance of suffering from malaria. Based on this estimate, what is the chance Guido's test result will indicate malaria?
- Given that his test result indicates malaria, what is the revised probability that Guido suffers from it?

Probability Example 2

- Hermione, who was born on January 24, enters a room with seven other people. Under the assumption of uniformly-distributed birthdays:
- What is the probability that none of the others shares her birthday?
- What is the chance that at least one of the others does so?

The Binomial Distribution

- Four basic properties:
  - Consists of n trials
  - Each trial has exactly two m.e. outcomes, A and B
  - The probability of A takes the same value, p, on all trials
  - The n trials are independent of each other

The Binomial Dist cont...

- X = number of times that event A comes up over the n trials of the binomial process
- $P(X=k) = \binom{n}{k} . p^k . (1 - p)^{n-k}$
- where $\binom{n}{k} = n! / (k!(n-k)!)$

Example

- Each sample of air has a 10% chance of containing a particular rare molecule. Assume the samples are independent with regard to the presence of the rare molecule. Find the probability that in the next 18 samples, exactly 2 contain the rare molecule.

The Mean (for discrete r.v.)

- a.k.a. the average, the expected value, the expectation
- For a discrete r.v.: $\mu = \Sigma x_i.P(x_i)$
  - Example: for binomial distribution
    $E(X) = np$

Rules for computing means

- $E(cX) = c.E(X)$
- $E(X+Y) = E(X) + E(Y)$
- $E(X) = E(X|A_1).P(A_1) + E(X|A_2).P(A_2) + \dots + E(X|A_m).P(A_m)$

Quick Example

- Determining the average revenue per summer day of a beachfront restaurant with average revenue of $33,000 on sunny summer days, and $14,000 on gloomy ones. If 70% of all days are sunny and the rest gloomy, what is the average revenue of the restaurant?

Variance and Standard Deviation

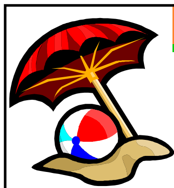
- For a discrete r.v.:

$Var(X) = \Sigma(x_i - \mu)^2.P(x_i)$

- Stddev (X) = $\sqrt{Var(X)}$

Variance Formulae

- $Var(c.X) = c^2.Var(X)$
- If X and Y are independent, then $Var(X+Y) = Var(X) + Var(Y)$
- Suppose that $E(Z|A_i) = \mu_i$ and $Stdev(Z|A_i) = \sigma_i$, then $Var(Z) = \Sigma\, [\sigma_i^2 + (\mu_i - E(Z))^2].P(A_i)$



Back to Restaurant Example

- Suppose now that $\sigma(Z|A_1)$ = \$4500 and $\sigma(Z|A_2)$ = \$1000. What is the Var(Z)?

Continuous Random Variables

- X is continuous -> X can assume any value in some interval from a to b.
- Probabilistic questions refer to *intervals* rather than specific values.
- The *probability density function*:

  $f(y)dy = P(y \leq X \leq y+dy)$
- The *cumulative distribution function*: $F(y) = P(X \leq y)$

The Uniform Distribution

- $X \sim U(a,b)$
- What is the *probability density function*?
- What is the *cumulative distribution function*?

Cdf and pdf

- $F(y) = P(X \leq y) = P(X < y)$
- $P(X > y) = 1 – P(X < y) = 1 – F(y)$
- $P(c \leq X \leq d) = F(d) – F(c)$

- How does the cdf relate to the pdf?
- Rewrite the above equations using the probability density function.

Example

- Hermione believes that the price of gas is equally likely between \$1.25 and \$1.60. Suppose four gas stations are chosen at random. Assuming that Hermione is correct, find the probability that:
- The first one chosen has price between \$1.25 and \$1.50.
- All four have prices between \$1.25 and \$1.50
- None have prices between \$1.25 and \$1.50
- At least one has price between \$1.25 and \$1.50.

The Normal Distribution

- X ~ N(μ,σ)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Using the Normal Table

- Compute $z = (y - \mu) / \sigma$
- From the normal table, $F(y) = \Phi(z)$
- That's all.

Some numbers of interest

- $P(x \le \mu) = P(x > \mu) = 0.5$
- $P(\mu - \sigma \le x \le \mu + \sigma) \approx 0.68$
- $P(\mu - 2\sigma \le x \le \mu + 2\sigma) \approx 0.95$
- $P(\mu - 3\sigma \le x \le \mu + 3\sigma) \approx 0.997$

Example: Earthquakes

- Some geological measurements suggest that maximum-strength earthquakes occur on the southern end of the San Andreas Fault every 145 years on average. But the individual intervals between “Big Ones” vary a bit around this average.

Earthquakes cont…

- Assume that an adequate model posits that, given that a huge earthquake has just occurred, the time x until the next one follows a normal distribution with mean 145 and standard deviation 10.
- The last huge earthquake on the southern San Andreas occurred in 1857. What is the probability that the next monstrous quake will take place before 2010?

Example: Computer Code

- Suppose that a computer program consists of 10,000 lines of code, each of which independently has a one in 1500 chance of containing an error. Guido, the dean of debugging, can detect and correct each error present in a program in (exactly) one hour. If he is hired to get the 10,000-line program ready to roll, what is the probability he will be able to do so in one eight-hour work day?

Normal approx to Binomial

- If X~Bin(n,p),
- If np>5 and n(1–p)>5,
- Then Z is approximately a standard normal random variable, where

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

Example: Coin Tossing

- If a fair coin is tossed 144 times, what is the probability that the number of heads falls between 60 and 72 (including the end points)?

Correlation

- Positive correlation
- Negative correlation
- Uncorrelated

Example

- Consider the quantities
  - Q = noon temperature in Bethlehem
  - R = beer consumption in Bethlehem that day
  - S = hot coffee consumption in Bethlehem that day
  - U = noon temperature in the Philadelphia

Coefficient of Correlation

- Definition:

$$\hat{\rho} = \frac{\mathrm{E}(XZ) - \mathrm{E}(X)\mathrm{E}(Z)}{\mathrm{s}(X)\mathrm{s}(Z)}$$

Example

- Suppose that, over a data set, x equals 1 half the time and 11 half the time, and z follows the same pattern. Find the correlation coefficient for the following cases:
- CASE I: whenever x=11, z=11, and whenever x=1, z=1
- CASE II: whenever x=11, z=1, and whenever x=1, z=11
- CASE III: when x=11, z is equally likely to be 1 or 11; when x=1, z is equally likely to be 1 or 11.

The Covariance

- Definition:

$$\mathrm{COV}(X, Z) = \mathrm{E}(XZ) - \mathrm{E}(X)\mathrm{E}(Z)$$