Boston University

Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Problem Set No. 1

Fall 2016

Issued: Wednesday, Sept. 7, 2016 **Due:** Monday, Sept. 19, 2016

Problem 1.1

A random experiment consists of tossing a die and observing the number of dots on the top face. Let A_1 be the event that 3 comes up, A_2 the event that an even number comes up, and A_3 the event that an odd number comes up.

(a) Find $P(A_1), P(A_1 \cap A_3)$.

(b) Find $P(A_2 \cup A_3)$, $P(A_2 \cap A_3)$, $P(A_1|A_3)$.

(c) Are A_2 and A_3 disjoint?

(d) Are A_2 and A_3 independent?

Problem 1.2

A group of students is taking a multiple-choice test. For a particular question on the test, the fraction of students who know the answer is p. The fraction that will have to guess the answer is $(1-p)$. If a student knows the answer, then he will certainly answer the question correctly. If a student doesn't know the answer and must guess, then the probability of answering the question correctly is $1/n$, where n is the number of choices for the given question.

(a) Compute the probability P_c that a student who answers the question correctly actually knew the answer.

(b) Suppose that the professor believes that $p = .85$, i.e. that 85% of the students actually know the answer. Further, suppose that he wants to design the multiple choice question such that $P_c = .95$, i.e. so that correct answers on the question indicate actual knowledge with 95% probability. How many parts n should the problem have?

Problem 1.3

You are a contestant on a game show. There are three closed doors leading to three rooms. Two of the rooms contain nothing, but the third contains a prize of \$ 100000 which is yours if you pick the right door. You are asked to pick a door by the compere who knows which room contains the car. After you pick a door, the compere opens a door (not the one you picked) to show an empty room. Show that, even without any further knowledge, you will greatly increase your chances of winning the car if you switch your choice from the door you originally picked. (Hint: Use Bayes theorem. By symmetry we can assume that you initially always picks the first door. The compere picks the second door. Define F_k as the event that the prize is behind door k. Define B as the event that the compere opens door 2 and there is no prize behind door 2.)

Problem 1.4

A random variable x has probability distribution function

$$P_X(x) = \left[1 - e^{-2x}\right] u(x)$$

where $u(\cdot)$ is the unit-step function.

(a) Calculate the following probabilities:

$$P[X \leq 1], \qquad P[X \geq 2], \qquad P[X = 2].$$

(b) Find $p_X(x)$, the probability density function for X.

(c) Let Y be a random variable obtained from X as follows:

$$Y = \begin{cases} 0, & \text{if } X < 2 \\ 1, & \text{if } X \geq 2 \end{cases}$$

Find $p_Y(y)$, the probability density function for Y.

Problem 1.5

Let X and Y be statistically independent random variables with probability density functions

$$p_X(x) = \frac{1}{2}\delta(x+1) + \frac{1}{2}\delta(x-1), \qquad p_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right)$$

and let $Z = X + Y$, and $W = XY$.

(a) Find the conditional probability density functions $p_{Z|X}(z|x = -1)$ and $p_{Z|X}(z|x = 1)$.

(b) Find the probability density function $p_Z(z)$ of Z.

(c) Find the mean values $m_X = E(X)$, $m_Y = E(Y)$, the variances, σ_Y^2, σ_W^2, and the covariance σ_{YW}. Are Y and W uncorrelated random variables? Are Y and W statistically independent random variables?

Problem 1.6

The outcome of a random experiment is known to have an exponential distribution, but the parameter α of that distribution is not known. We estimate the parameter α from the sample mean:

$$\widehat{\alpha} = \frac{1}{\widehat{m}} \qquad \widehat{m} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

where x_i are independent trials of the experiment.

(a) Note that $\widehat{m}$ is itself a random variable. Find the mean and variance of $\widehat{m}$.

(b) Use the Chebychev bound to estimate the minimum number of experiments N that are required to guarantee that

$$P[|\widehat{m} - m| > .01m] \leq .001$$

where $m = 1/\alpha$.

(c) Use the Central Limit Theorem to approximate the distribution of the random variable $\widehat{m} - m$ as a Gaussian random variable. Using the Gaussian approximation, estimate the minimum number of experiments N which would be required to guarantee that

$$P[|\widehat{m} - m| > .01m] \leq .001$$

Note: You will need to find a table of values of the "Q" function: $Q(\beta) = \int_\beta^\infty \frac{1}{\sqrt{2\pi}} e^{y^2/2}\, dy$.

(d) (Bonus) Use Chernoff's bound(Page 36 in Lecture Notes) to estimate the minimum number of experiments N above.

Problem 1.7

In the mid to late 1980's, in response to the growing AIDS crisis and the emergence of new, highly sensitive tests for the virus, there were a number of calls for widespread public screening for the disease. The focus at the time was the sensitivity and specificity (roughly, 1-false positive rate) of the tests at hand. For the tests in question the sensitivity was Pr(Positive Test | Infected) ≈ 1 and the false positive rate was Pr(Positive Test | Uninfected) $\approx .00005$ – an unusually low false positive rate. What was generally neglected in the debate, however, was the low prevalence of the disease in the general population: Pr(Infected) ≈ 0.0001. Since being told you are HIV positive has dramatic ramifications, what clearly matters to you as an individual is the probability that you are uninfected given a positive test result: Pr(Uninfected | Positive test). Calculate this probability. Would you volunteer for such screening? How does this number change if you are in a "high risk" population – i.e. if Pr(Infected) is significantly higher?

References

- K. B. Meyer and S. G. Pauker, "Screening for HIV: Can we afford the False Positive Rate," The New England Journal of Medicine, Vol 317, No 4, pg 238–241, 1987.
- R. Weiss and S. O. Thier, "HIV Testing is the Answer – Whats the Question?," The New England Journal of Medicine, Vol 319, No 15, pg 1010–1012, 1988.
- J. B. Jackson et al, "Absence of HIV Infection in Blood Donors with Indeterminate Wester Blot Tests for Antibody to HIV-1," The New England Journal of Medicine, Vol 322, No 4, pg 217–222, 1990.

Problem 1.8
Suppose $\underline{X}$ is a Gaussian random vector with mean vector and covariance matrix specified below. Find the mean vector and covariance matrix of $\underline{Y} = [Y_1, Y_2, Y_3]^T$, where Y_1, Y_2, and Y_3 are specified below.

$$m_{\underline{X}} = \begin{bmatrix} 6 \\ 0 \\ 8 \end{bmatrix}, \quad \Lambda_{\underline{X}} = \begin{bmatrix} 1/2 & 1/4 & 1/3 \\ 1/4 & 2 & 2/3 \\ 1/3 & 2/3 & 1 \end{bmatrix}, \quad \begin{aligned} Y_1 &= X_1 - X_2 \\ Y_2 &= X_1 + X_2 - 2X_3 \\ Y_3 &= X_1 + X_3 \end{aligned}$$

Computer Projects

In the course we will focus on *models* of stochastic phenomona. These models are useful abstractions of reality. In this set of projects we will begin to connect our models to data. The class web site contains two functions, `pdf1d.m`, and `pdf2d.m` which generate estimates of the probability density function (pdf) and probability distribution function (PDF/CDF) of a single and 2 joint random variables, respectively, given vectors of independent observations of these random variables. The estimate is based on a scaled histogram of the data (how is a pdf estimate different from a simple histogram?). You will use these functions in the following problems.

Problem 1.9 Transformations of Random Variables
The purpose of this problem is to investigate transformations of random variables. One very important practical task is the computer generation of random variables with a *given* distribution. In this project we will see how this can be done.

(a) Let X be a uniformly distributed random variable and define the transformed random variable $Y = g(X)$, where the mapping $g(\cdot) : [0,1] \mapsto [-\infty, \infty]$ (i.e. $g(\cdot)$ maps the interval $[0,1]$ to the entire y axis). Suppose that $g(\cdot)$ is also a monotonically increasing function of X, so that there is a unique X corresponding to each Y. What is $P_Y(y)$ in terms of the function $g(\cdot)$? (Note: this solves the analysis or forward problem, in that if a $g(\cdot)$ is specified it explains what the transformed PDF is).

Use this answer to specify a way to generate a random variable Y with a *given* PDF/CDF $P_Y(y)$ starting from a uniformly distributed random variable X. (Note: This solves the design problem, in that the target PDF is given and an appropriate function $g(\cdot)$ to achieve it is sought).

(b) An important random variable is the Cauchy random variable. This random variable (and the Laplace random variable) is often used as a model of impulsive noise in systems. Such distributions are sometimes termed "heavy tailed", referring to the sizeable probability mass in the tails of the distribution (i.e. far from the peak). Write a program `randcau.m` to generate Cauchy random variables. Generate and plot a few thousand Cauchy random variables to see the impulsive behavior. How does the heavy tailed nature of the distribution exhibit itself when trying to generate a pdf estimate using e.g. `pdf1d.m`?

(c) The above method only works when an analytic expression for the inverse of $P_Y(y)$ can be found. One important situation where this is not the case is the Gaussian random variable. Suggest different ways you might generate (approximations to) zero mean, unit variance Gaussian random variables from uniform random variables. Try out your schemes and verify the results using `pdf1d.m`.

Problem 1.10 Joint Probability Density Functions, Conditionals, and Marginals
In this problem we will investigate joint random variables and their properties.

(a) On the class web site there is a data set `2rvdata.mat` with 4 different sets of joint random variables pairs `(xi,yi)`. Our goal is to understand these random variable pairs better. For each of these sets of joint random variables perform the following calculations:

   (i) Generate and plot the empirical joint pdf $p_{X_i,Y_i}(x_i, y_i)$ and PDF $P_{X_i,Y_i}(x_i, y_i)$.
   (ii) Generate and plot the empirical marginal distributions $p_{X_i}(x_i)$ and $p_{Y_i}(y_i)$.
   (iii) Calculate the empricial covariance matrix and the corresponding correlation coefficient. For your estimate of the covariance matrix approximate the expection operation with the sample mean.
   (iv) Estimate and plot the empirical conditional pdf at e.g. $p_{X_i|Y_i}(x_i|y_i = .5)$ and $p_{X_i|Y_i}(x_i|y_i = .75)$.

(b) Based on the information in (a), specify which of the data sets you think came from uncorrelated random variables and which came from independent random variables. For example, you may want to compare $p_{X_i,Y_i}(x_i, y_i)$ to $p_{X_i}(x_i)p_{Y_i}(y_i)$ when deciding independence, etc. Explain your reasoning.