Probability and Statistics
TD n^o 1

October 17, 2012

1 Axioms of probability

Prove that the conditional probability $P(.|F)$ (F is a fixed event) satistifies the three axioms of probability.

2 Independence

1. Prove that if E and F are independent, then so are E and F^c

2. Two fair dice are thrown. Let E_1 denote the event that the sum of the two dice is 6 and E_2 denote the event that the sum of the two dice is 7. Let also F denote the event that the first die equals 4 and G the event is that the second die equals 3.

   (a) Prove that F (or G) is independent of E_2 but not of E_1

   (b) Illustrate with F,G and E_2 that while one event might be independent of two other events, it is not necessarily independent from their intersection

3 $E[1/X]$ vs. $1/E[X]$

In a real observation of a peer-to-peer file replication application, i.e an application that aims at replicating a given file on a large number of peers as fast as possible, it has been observed that:

- The mean download throughput t_0 of clients was 1Mb/s;
- The mean download time d_0 per client was 30,000 s;

The file to be replicated was of size $s = 2GB$.

1. What would have been the download time d_1 if all clients had experienced the same download rate?

2. What would have been the download time d_2 if half of the clients experience a rate of $500kb/s$ and half of the clients experience a rate of $1.5Mb/s$?

3. Formulate an empirical law relating d_0 to (some property) of the random variable that denotes the download throughout T experienced by clients

4 Conditional Probabilities

A factory receives hard-disks from two different manufacturers, M_A and M_B. The probability that the piece comes from M_A (resp. M_B) is $P(M_A) = 0.6$ (resp. $P(M_B) = 0.4$). From prior experience, we know that 2% of the disks of M_A are likely to fail while it is 6% for M_B. If we observe a failure, what is the probability that it comes from manufacturer M_A?

5 Expectation, Variance, Quantiles

1. Consider a discrete random variable X that takes values $x_i; i \geq 1$ with probabilities $p(x_i)$. Prove that:
$$E[g(X)] = \sum_i g(x_i)p(x_i) \tag{1}$$

2. Compute the Variance of $Y = aX + b$.

3. Let $x_1, x_2, x, ...x_{100}$ be 100 samples following the same distribution. What is the value of the 40% quantile of these samples?

4. Let X and Y be discrete random variables.
   (a) Prove that $E[X + Y] = E[X] + E[Y]$
   (b) Compute and give an interpretation to $V[aX + b]$ where a and b are scalar

5. Consider a non-negative continuous random variable X. Prove the following formula (this formula is useful in practice to compute $E[X]$).
$$E[X] = \int_{x=0}^{+\infty} \bar{F}(x)dx, \qquad \text{where } \bar{F}(x) = 1 - F(x) = P\{X > x\} \tag{2}$$

6 Sequence of trials

An infinite sequence of trials is performed. Each trial results in a success with probability p and a failure with probability 1-p. What is the probability that:

1. at least one success occurs in the first n trials;
2. exactly k successes occur in the first n trials;
3. all trials result in a success (look for an intuitive proof);

7 Exponential pdf

The lifetime in years of a flat panel display is a random variable that follows the exponential pdf given by:

$$f(x) = 0.1e^{-0.1x} \tag{3}$$

1. What is the mean lifetime of the flat panel display?

2. What is the probability that the flat panel fails within the first two years?

3. Given that the display has been in operation for one year, what is the probability that it will fail in the next year?

8 Poisson Distribution

1. Compute $E[X]$ and $V[X]$ when X is a Poisson random variable

2. Compute the density of two independent Poisson random variables

3. Let us suppose that events are occurring at certain random points in time, and let us assume that there exists a constant λ such that:

    - $P(1 \text{ event in } [t, t+h]) = \lambda h + o(h)$ for all $t \geq 0$
    - $P(2 \text{ or more events in}[t, t+h]) = o(h)$ for all $t \geq 0$
    - For any integer n, $j_1, ...j_n$ and n non-overlapping intervals, if E_i is the event that exacly j_i events occur during interval i, then $E_1, ...E_n$ are independent.

    The objective is to prove that the number of events $N(t)$ that fall in the interval $[0, t]$ is a Poisson random variable with parameter λt. The starting point is to break the interval $[0, t]$ into n sub-intervals of equal size. Next, to evaluate $P\{N(t) = k\}$, express the event $\{N(t) = k\}$ by conditioning on the fact that either k points fall in k distinct sub-intervals or at least one sub-interval has 2 or more points. Then, evaluate each of the two probabilities separately.

    it follows by the same argument that verified the Poisson approximation to the binomial approximation (see course notes) that when $n \to \infty$:

    $$P\{N(t) = k\} = e^{\lambda t} \frac{(\lambda t)^k}{k!} \tag{4}$$

9 Gamma Distribution

Considering the events randomly occuring in time of exercise 8, prove that the amount of time T_n that one has to wait until a total of n events occur follows a gamma distribution with parameter n, λ.

10 Random Variables and their Realizations

1. Use the distool and randtool for a few distributions (exponential,normal, Poisson) to check the possible discrepencies between the shapes of the densities and the histograms of random samples.
2. Check that as the number of samples increases, the accuracy (fitting) is better, though there are still outliers
3. Consider the normal distribution. Observe how the maximum of the pdf varies when μ or σ varies.

11 Laplacian probability density function

When a random variable is equally likely to be either positive or negative, then the Laplacian distribution can be used to model it. The Laplacian pdf with parameter $\lambda > 0$ is given by:

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|} \tag{5}$$

1. Derive the mean, the variance and the cumulative distribution function for the Laplacian.
2. Write a matlab function that evaluates the Laplacian cdf and plot it for different values of λ. (use the quad function of matlab)

12 Poisson Approximation

Compare the Poisson approximation to the actual binomial probability $P(X = 4)$ for n=10,100,... and p=0.0001, 0.01,... using matlab.

13 Computing the Binomial Distribution Function

Consider a binomial variable X. Compute the ratio $\frac{P(X=i)}{P(X=i-1)}$ and based on it, develop a simple a program that implement the computation of the density and the cdf of X in matlab. Check the result against the functions implemented in the statistical toolbox of matlab, binopdf and binocdf.