ECSE-4690: Experimental Networking

Informal Quiz: Prob/Stat

Shiv Kalyanaraman: shivkuma@ecse.rpi.edu

Rensselaer

RSVP

Probability/Statistics

- A probability density function (PDF) is a generalization of a histogram for the continuous random variable case.
- A random variable (R.v.) models a measurement, whereas probability models an experiment, and r.v. is used when the measurement does not necessarily captures the set of all possible outcomes of the experiment.
- In the experiment of tossing a die, the set X = {0,1,2} which denotes the possibility of the outcomes being 0, 1 or 2 is a random variable.
- A mean of a random variable is also known as the first moment or centroid of a distribution.
- A median is the 50^{th} percentile element, found using the inverse of the CDF with an argument of 0.5.
- A mean is the preferred central tendency measure in a skewed distribution.
- A mode (or the most probable element) is usually used with categorical random variables instead of mean or median
- C.o.V. and SIQR are measures of central tendency.
- Covariance, a measure of dependence between random variables, always lies between –1 and +1

Probability/Statistics

- If E(XY) = E(X)E(Y), the random variables X and Y are independent
- Coefficient of Variation (C.o.V) and Correlation Coefficient (ρ_{XY}) are normalized measures of spread and dependence respectively.
- The C.o.V would be a useful metric to measure the unfairness of rate allocations to TCP flows passing through a single bottleneck
- The correlation coefficient would be a useful metric to measure the degree of traffic and window synchronization between a pair of TCP flows competing at a bottleneck
- Given 50 RTT samples, one can estimate the 95% confidence interval of the path RTT and a good estimate of maximum RTT (to set the timeout value in TCP)
- A Bernoulli distribution can be studied by considering a sequence of N bernoulli trials, and counting the number of successes in N trials.
- Taking a large bet with a probability of success 0.5 in a single experiment (like a lottery, without regard to cost) is superior to taking smaller bets (with probability 0.01 each) in 50 repeated, identical experiments. (Hint: probability of success in latter case is $1 - (0.99)^{50}$)
- The Poisson distribution is a continuous-time approximation of the binomial distribution, derived by assuming $np = \lambda$, and n is very large.
- In a Poisson arrival process, the average time since the occurrence of the last arrival is the same as the average time for the next arrival.

Probability/Statistics

- ❑ ❑ The Chebyshev bound for spread of a random variable is a very loose bound, especially for the normal distribution.
- ❑ ❑ The distribution of sample means from any distribution (I.e. sampling distribution, assuming random sampling) tends to a normal distribution
- ❑ ❑ Confidence interval gives less information compared to the notion of "statistical significance" and "null hypothesis"
- ❑ ❑ A t-distribution is an approximation of the normal distribution with n-1 degrees of freedom that can be constructed with n samples from a normal population & the approximation is good when n is at least six.
- ❑ ❑ The confidence interval is constructed from a normal or normal-like distribution (eg: t-distribution) of a random variable (eg: the sample mean) by excluding the tails of the distribution based upon the given confidence level
- ❑ ❑ Pairing and randomized experiments are ways of ensuring the random sampling assumption and reducing correlations between experiments
- ❑ ❑ If two confidence intervals for an estimate of a mean overlap and the means also lie in the CIs of each other, the means cannot be declared to be different at that level of confidence.