CS109 Final Review Problems

June 1, 2016

Some CDFs

Compute the CDF of the following distributions

(a) The *uniform distribution*

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & else \end{cases}$$

(b) The *Rayleigh Distribution*

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} & x \geq 0 \\ 0 & else \end{cases}$$

This gives the probability density of the length of a vector in $\mathbb{R}^2$ which has components which are independent zero mean Gaussian random variables.

(c) The standard *Cauchy Distribution*

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

This gives the probability density of the ratio of two standard normal random variables.

Functions of a Random Variable

Let U be the random variable which is uniformly distributed on (0,1]. Compute the distributions of the following random variables.

(a) $X = U^{1/2}$

(b) $Y = -\ln U$

(c) $Z = aU + b \quad a < 0, b > 0$

Chi Squared Distribution

From years of extensive data collection, you and your colleagues determine the length of adult snakes in a nature preserve are well modeled as a Gaussian random variable. Specifically, suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ where X is the length of the snake. Since you have a good model for size of the snakes, you want to characterize the deviation of a snake's length from the mean to determine if a snake is abnormally large or small for the preserve. A large number of snakes that deviate "too much" from the mean might indicate the model is no longer correct and perhaps an underlying enviromental change causing it.

You and your colleagues are agnostic to whether the snakes are too far below or above the mean, since any deviation might indicate a failing of the model. You propose a good measurement to characterize this is the random variable $Y = (X - \mu)^2/\sigma^2$. In order to calculate the probability a snake deviates by a certain amount from the mean, we need a distribution for Y. Find the PDF of y.[1]

Discretizing a Continous RV

A large call center has determined that the number of minutes until an employee is able to serve the next customer on hold is distributed $X \sim \text{Exp}(\lambda)$. The amount of time it takes to answer a customer on hold directly affects the call center's revenue as well as their customer's satisfaction. As a result the call center naturally wants to do some analysis on the time taken to serve its customers.

The company decides that the difference between waiting 5 minutes versus waiting, say 5 minutes and 23 seconds, is more or less irrelevant and they would instead like to count them both as having taken 5 minutes. Rather than model X which can take on a continous range of values, they want a model for $Y = \lfloor X \rfloor$.

Find the distribution of this random variable and its relevant parameters. Though wrong, an answer might be something like $\text{Poi}(\log \lambda)$.

Windfarm Modeling

Wind velocity can be expressed in terms of its north and east components denoted v_x and v_y respectively. On our wind farm from the midterm, the north and east components of wind velocity can be modeled as independent gaussian random variables with distribution $v_x, v_y \sim \mathcal{N}(0, \sigma^2)$ during the summer months. The overall magnitude of the wind is then distributed as a *Rayliegh Distribution* given by the PDF

$$f_X(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} & x \geq 0 \\ 0 & else \end{cases}$$

Always interested in evaluating the effectiveness of our business, we wish to model the wind speed on our farm. To this end we collect N independent measurements of wind speeds (the magnitude of velocity) $w_1, w_2, \cdots, w_N$. Find a maximum likelihood estimate of σ^2 if we are modeling the wind speed as coming from a Rayleigh distribution

[1] This random variable is distributed chi squared with one degree of freedom. Roughly, it gives the distribution of how far away from the mean a gaussian random variable will be (normalized by the variance)