The Calculus of Statistics

Blaise Whitesell Jeremiah Gelb

May 25, 2015

1 Probability Distribution Functions

1.1 Probability Density Functions

A probability density function (**PDF**) describes the relative likelihood $f(x)$ of possible outcomes for a continuous random variable.

- The probability of any x occurring is either positive or zero, so a PDF can never be negative.

$$f(x) \geq 0 \quad \text{for all } x$$

- The area under the curve must equal 1

$$\int_a^b f(x)\,\mathrm{d}x = 1$$

where a and b are the lower and upper bounds, often $-\infty$ and ∞.

1.2 Cumulative Distribution Functions

A cumulative distribution function (**CDF**) gives the probability $F(x)$ that the outcome of a continuous random variable will be less than or equal to x.

- The CDF is related to the PDF by this integral:

$$F(x) = \int_a^x f(t)\,\mathrm{d}t \quad \text{for } a \leq x \leq b$$

- A CDF is monotonically increasing on the interval (a, b).

$$F(a) = 0 \qquad F(b) = 1$$

1.3 Using Calculus

To find the probability of an outcome in a certain range, one can integrate the PDF over that interval (c, d) contained in (a, b).

$$\int_c^d f(x)\,\mathrm{d}x \quad = F(d) - F(c)$$

This gives the area under the curve, corresponding to the probability.

2 The Uniform Distribution

2.1 Definition

A uniform distribution describes a continuous random variable where every outcome on an interval (a, b) is equally likely.

$$f(x) = \kappa$$

2.2 Implications

- The distribution looks like a rectangle with length $b - a$ and height κ
- Area $= 1 = \kappa(b - a)$

$$\int_a^b \frac{1}{b-a}\,\mathrm{d}x = \left.\frac{x}{b-a}\right|_a^b = \frac{b}{b-a} - \frac{a}{b-a} = \frac{b-a}{b-a} = 1$$

- Therefore, we have an explicit definition for κ:

$$\kappa = \frac{1}{b-a}$$

2.3 PDF of a uniform distribution

The formal equation is piecewise. It depends only on the bounds a and b.

$$f(x) = \begin{cases} 0 & \text{if } x < a, \\ 1/(b-a) & \text{if } a \leq x \leq b, \\ 0 & \text{if } x > b. \end{cases}$$

2.4 CDF of a uniform distribution

The CDF increases linearly on the interval (a, b).

$$f(x) = \begin{cases} 0 & \text{if } x < a, \\ (x-a)/(b-a) & \text{if } a \leq x \leq b, \\ 1 & \text{if } x > b. \end{cases}$$

Problem 1. In the game "Spin to Win", the player spins a giant roulette wheel with 200 spaces. One is marked "Win", one is marked "Lose", and the remaining 198 are marked "Spin Again". What is the probability of losing on the first spin?

Solution. There are 200 spaces, exactly one of which corresponds to losing immediately. The continuous range of values can be modeled as a uniform distribution from $0°$ to $360°$.

$$1/200 \times 360° = 1.8°$$

$$\int_0^{1.8} \frac{1}{360}\,\mathrm{d}\theta = \left.\frac{\theta}{360}\right|_0^{1.8} = \frac{1.8}{360} - \frac{0}{360} = 0.005$$

Interestingly enough, this is equivalent to 1/200. □

---

3 The Exponential Distribution

3.1 Introduction

An exponential distribution can be used to model events that occur independently at a constant average rate.

Problem 2. 250 kids at a frat party are randomly getting sick at a continuous rate of 20% per hour, beginning at 12:00AM. What are the odds that Chad vomits between 2:00AM and 3:00AM?

Solution. Let $t = 0$ at 12:00AM. Then the number of kids who are not sick is given by $250e^{-.2t}$. This means that the number of kids who are sick is given by $250 - 250e^{-.2t}$. Therefore, the proportion of kids who are sick after t hours is

$$\frac{250 - 250e^{-.2t}}{250} = 1 - e^{-.2t}$$

This is equivalent to the probability of being sick after t hours, which is the CDF. To find the PDF, simply take a derivative.

$$\begin{aligned} \int_0^t f(x)\,\mathrm{d}x &= 1 - e^{-.2t} \\ \frac{\mathrm{d}}{\mathrm{d}t}\int_0^t f(x)\,\mathrm{d}x &= \frac{\mathrm{d}}{\mathrm{d}t}(1 - e^{-.2t}) \\ f(x) &= .2e^{-.2t} \end{aligned}$$

Now that we know the PDF, we integrate it between $t = 2$ and $t = 3$.

$$\int_2^3 .2e^{-.2t} = -e^{-.2t}\Big|_2^3 = -e^{-.6} - (-e^{-.4}) = .1215$$

Note that we could also have found this answer by finding the difference between $F(3)$ and $F(2)$, without using the PDF at all.

$$\left(1 - e^{-.2t}\Big|_{t=3}\right) - \left(1 - e^{-.2t}\Big|_{t=2}\right) = .1215$$

The probability that Chad will vomit between 2:00AM and 3:00AM is .1215. $\square$

---

4 The Beta Distribution

5 The Normal Distribution