Formula Sheet – Final Project

Christian Simpson

CSCI-3327-001 Probability and Applied Statistics

Byron Hoy

April 29th, 2024

No table of contents entries found.

Poisson Distribution

The Poisson probability distribution often provides a good model for the probability distribution of the number Y of rare events that occur in space, time, volume, or any other dimension, where λ is the average value of Y.

The mass function representing a variable with poisson distribution is given by

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}$$

Where y is the number of occurrences of an event, and lambda (λ) represents both the expected $E(Y)$ and the variance $V(Y)$.

The random variable y must be continuous on the interval $0 \leq y \leq n$ with λ > 0.

Tchebysheff's Theorem

In many instances, the shapes of probability histograms differ markedly from a mound shape, and the empirical rule may not yield useful approximations to the probabilities of interest. The following result, known as Tchebysheff’s theorem, can be used to determine a lower bound for the probability that the random variable Y of interest falls in an interval μ ± kσ.

Let Y be a random variable with mean μ and finite variance σ^2. Then, for any constant k > 0,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} \text{ or } P(|Y - \mu| \geq k\sigma) \leq 1 - \frac{1}{k^2}$$

Continuous Random Variables

A random variable that can take on any value in an interval.

i.e., inches of rainfall in an area; years of operation of a washing machine.

Differs from discrete in that the random variable Y can take on any of the infinite values on an interval instead of discrete, countable numbers.

Probability Distribution Functions

Distribution Function: $F(y) = P(Y \leq y)\ on\ -\infty < y < \infty$

"For some random variable Y, the distribution function F of y is denoted by the Probability of that random variable taking on some value less than or equal to y on the interval between negative infinity and infinity."

The **Cumulative Distribution Function** for a variable denotes the sum of probabilities for some value the variable may take on for all values at or before the current value.

Example: for some continuous random variable Y with the following probabilities:

$$p(0) = \frac{1}{4} p(1) = \frac{1}{2} p(2) = \frac{1}{4}$$

The distribution function $F(y)$ is displayed as:

$$0\ for\ y < 0$$

$$F(y) = \frac{1}{4} for\ 0 \leq y < 1$$

$$\frac{3}{4} for\ 1 \leq y < 2$$

$$1\ for\ y > 2$$

Properties of a Distribution Function

1. The limit as y approaches negative infinity is zero. $\lim_{y \to -\infty} F(y) = 0$
2. The limit as y approaches infinity is one. $\lim_{y \to \infty} F(y) = 1$
3. The function must be non-decreasing. $F(y_1) \leq F(y_2)$

Probability Density Functions

Density Function: $f(y) = F'(y)$

The probability density function is a theoretical model for the frequency distribution (histogram) of a population of measurements.

A probability density function for a random variable Y can be represented as the first derivative of that variable's distribution function.

If the random variable Y has density function f (y) and a < b, then the probability that Y falls in the interval [a, b] is given by the equation

$$P(a \leq Y \leq b) = \int_a^b f(y)\ dy$$

Expected and Variance of Continuous Random Variables

Expected for a C.R.V.:

The expected (mean) of a continuous random variable Y is given by

$$E(Y) = \int_{-\infty}^{\infty} y \, f(y) \, dy$$

Example:

If a density function for Y is given as $f(y) = \frac{3}{8}y^2$ for $0 \le y \le 2$, then its expected value can be calculated as $E(Y) = \int_0^2 y \left(\frac{3}{8}\right) y^2 \, dy \rightarrow \left(\frac{3}{8}\right)\left(\frac{y^4}{4}\right)$ on $0 \le y \le 4$ = 1.5.

Expected for a <u>function</u>:

The expected for a function $f(y)$ is given by

$$E\big(g(y)\big) = \int_{-\infty}^{\infty} g(y) \, f(y) \, dy$$

Variance for a C.R.V.:

The variance of a continuous random variable Y is given by

$$V(Y) = E(Y^2) - E(Y)^2$$

Where $E(Y^2)$ is given subsequently by

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 \, f(y) \, dy$$

In the case of calculating the variance, $E(Y)^2$ represents squaring the expected.

Uniform Distribution

A continuous probability distribution in which the random variable takes on values such that $0 \leq y \leq 1$.

The density function for a continuous random variable with Uniform Distribution is given as

$$f(y) = \frac{1}{b-a}\ on\, a \leq y \leq b$$

The probability the random variable Y falls on an interval $c \leq y \leq d$ is given by

$$P(c \leq y \leq d) = \int_c^d \frac{1}{b-a}\, dy$$

Or simply

$$\frac{d-c}{b-a}$$

Expected: $E(Y) = \frac{b+a}{2}$

Variance: $V(Y) = \frac{(b+a)^2}{12}$

NOTE: For any interval between a and b, when considering intervals of equal range, each probability for those intervals should be the same. Thus, the name "Uniform Distribution".

Gamma Distribution

Some random variables are always nonnegative, yielding distributions of data that are skewed to the right – meaning most of the area under the curve is located near the origin, and the density function drops gradually as Y increases. The populations associated with these random variables frequently possess density functions that are adequately modeled by a gamma density function.

Examples: time between malfunctions of an aircraft engine; time between arrivals at a checkout counter; time to complete a maintenance check on a car.

These distributions have 2 distinct parameters:

α “alpha” = the shape parameter: determines shape of the curve.

β “beta” = the scale parameter: determines the magnitude of the scale.

The gamma density function in which α = 1 is called the exponential density function, given by

$$f(y) = \frac{1}{\beta} e^{-\frac{y}{\beta}}$$

Which exists on the interval $0 \leq y < \infty$.

Expected for Gamma Distribution: $E(Y) = \alpha\beta$

Variance for Gamma Distribution: $V(Y) = \alpha\beta^2$

Reminders for Double Integrals

For an area represented by the double integral

$$\int_a^b \int_c^d xy \, dA$$

Where x exists on the interval $0 \leq x \leq 2$ and y on the interval $0 \leq y \leq 1$, we can evaluate each integral separately. We do this by choosing a variable to be the inner integral and a variable for the outer. These values can be either variable as long as the intervals respect the following rules:

1. The outer limits must be constant.
2. The inner limits may be dependent.

Evaluating double integrals is simple, provided the integration is done in the correct order.

The inner integral must always be evaluated first, and the outer second. Each variable is integrated with respect to the other.

For the above example, if x is the inner variable:

$$\int_0^1 \int_0^2 xy \, dx \, dy$$

We integrate x on its interval first, treating y as a constant:

$\int_0^1 y \frac{x^2}{2} \, dy$ on $0 \leq x \leq 2$

Calculating this interval leaves us with the equation

$$\int_0^1 4y \, dy$$

Which simply integrates to 2.

No matter which order you place the variables in, if their intervals follow the rules, the double integral should always integrate to the same value.

Multivariate Probability Distributions

As described, concerned with multiple random variables. Can have any number of random variables associated with a single function.

Bivariate and Multivariate Probability Distributions

The mass function for a Multivariate Probability Distribution is given by

$$p(y_1,\ y_2,\ \ldots,\ y_n)$$

Which can be interpreted as “the joint probability of the variables $y_1,\ y_2,\ \ldots,\ y_n$ to take on values where $Y_1 = y_1\ \cap\ Y_2 = y_2\ \cap\ \ldots$

The distribution function for joint multivariate functions is given by

$$F(y_1,\ y_2)\ =\ \int_{-\infty}^{y_1}\int_{-\infty}^{y_2} f(t_1,\ t_2)dt_2\ dt_1$$

The joint density function for random variables is given by

$$f(y_1,\ y_2)$$

Marginal and Conditional Probabilities

Let Y_1 and Y_2 be jointly discrete random variables with probability function $p(y_1,\ y_2)$. Then the **Marginal Probability Functions** of Y_1 and Y_2 are given by

$$p_1(y_1)\ =\ \sum_{all\ y_2}^{p}(y_1,\ y_2) \text{ and } p_2(y_2)\ =\ \sum_{all\ y_1}^{p}(y_1,\ y_2)$$

Which can be interpreted as “The probability of Y1 to take on some value y for all values of y2”. And vice-versa.

Let Y1 and Y2 be jointly continuous random variables with joint density function f (y1, y2). Then the marginal density functions of Y1 and Y2 are given by

$$f_1(y_1)\ =\ \int_{-\infty}^{\infty} f(y_1,\ y_2)\ dy_2 \text{ and } f_2(y_2)\ =\ \int_{-\infty}^{\infty} f(y_1,\ y_2)\ dy_1$$

For each density function, the approach is to integrate with respect to a the opposing variable when determining where the function exists.

Let Y1 and Y2 be jointly continuous random variables with joint density $f(y_1,\ y_2)$ and marginal densities $f_1(y_1)$ and $f_2(y_2)$, respectively. For any y_2 such that $f_2(y_2)\ >\ 0$, the conditional density of Y_1 given $Y_2\ =\ y_2$, and for any y_1 such that $f_1(y_1)\ >\ 0$, the conditional density of Y_2 given $Y_1\ =\ y_1$, are given by

$$f(y_1 \nmid y_2)\ =\ \frac{f(y_1, y_2)}{f_2(y_2)} \text{ and } f(y_2 \nmid y_1)\ =\ \frac{f(y_1, y_2)}{f_1(y_1)}$$

Independent Random Variables

Two variables x and y with a joint distribution function $F(x, y)$ are said to be independent if

$$F(x, y) = F_1(x) F_2(y)$$

That is, if the joint distribution function equals the product of the two individual distribution functions.

For example, consider the dice tossing probabilities.

$$p(1, 2) = \frac{1}{36} \text{ and } p_1(1) = \frac{1}{6} \cdot p_2(2) = \frac{1}{6}$$

Therefore, the variables are independent.

The same can also apply to joint density functions:

$$f(x, y) = g(x)h(y)$$

If the joint density function equals the product of the two individual functions, the variables are independent.