Random Variable and Probability Distribution

Random Variables

- A *random Variable* is a function from a sample space into the set of real numbers.
- Example: A coin is tossed twice. The sample space is $S = \{HH, HT, TH, TT\}$. Denote the sample points as $e_1 = HH$, $e_2 = HT$, $e_3 = TH$, $e_4 = TT$.

  – Let Y be the number of heads in two tosses. Then Y is a random variable. $Y(e_1) = 2, Y(e_2) = 1$, $Y(e_3) = 1$, $Y(e_4) = 0$.

  – Let X be the number of tails − number of heads. Then X is another random variable. $X(e_1) = -2$, $X(e_2) = 0$, $X(e_3) = 0$, $X(e_4) = 2$.

- A random variable partitions the sample space into disjoint subsets on each of which the value of the random variable is constant.

Discrete Random Variables and Its Probability Distribution

- A random variable is *discrete* if it assumes only a finite or countably infinite number of values.
- The probability that a random variable Y takes the value y is the probability of the set of sample points e_i for which $Y(e_i) = y$. This probability is denoted by $P(Y = y)$ or $p(y)$.
- The *probability distribution of a discrete random variable* is a list of values or function $p(y)$ for all y in the range of Y where $p(y)$ is as defined above. This $p(y)$ is also called the probability mass function (p.m.f.) of the random variable Y. Note that

  the book uses the term "probability function" instead of probability mass function.

- $\sum_y P(Y = y) = 1$, where the sum is taken over all possible values that Y can assume.
- Example: Let Y be the number of heads in tow tosses of a balanced coin. Then Y can take values 0,1 or 2 and the distribution of Y is given by $p(0) = \frac{1}{4}$, $p(1) = \frac{1}{2}$ and $p(2) = \frac{1}{4}$.

Probability Histogram

- The probability distribution of a discrete random variable can be represented by a probability histogram for simple cases.
- A probability histogram is similar to a regular histogram except that the area represents probability instead of frequency.

Continuous Random Variables

- A continuous random variable can assume an uncountably infinite number of values.
- The probability distribution of a continuous random variable can not be described by the probabilities $P(Y = y)$.

Cumulative Distribution Function (c.d.f.)

- The cumulative distribution function or distribution function of a random variable Y, denoted by $F(y)$ is defined as

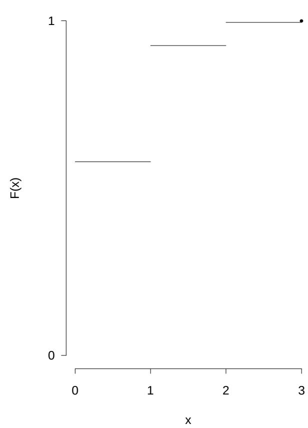
$$F(y) = P(Y \leq y) \qquad \text{for } -\infty < y < \infty$$

- For a discrete random variable the c.d.f. is a step function.

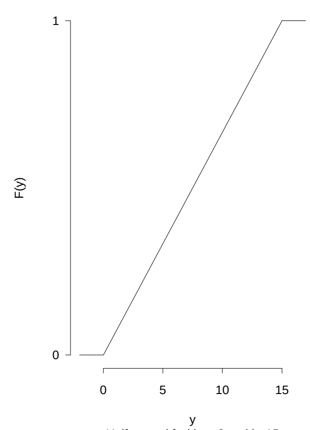
Properties of c.d.f.

1. $F(-\infty) = \lim_{y \to -\infty} F(y) = 0$
2. $F(\infty) = \lim_{y \to \infty} F(y) = 1$
3. $F(y)$ is a nondecreasing function, i.e., if $y_1 < y_2$, then $F(y_1) \leq F(y_2)$.
4. $F(y)$ is right continuous.

Examples of c.d.f. in Discrete and Continuous Case.



Binomial c.d.f with n=3, p=1/6



Uniform c.d.f with a=0 and b=15

Continuous Random Variable and Density Function

- A random variable is said to continuous if its c.d.f. is continuous.
- If a random variable is continuous with $F(y)$ as its c.d.f., it has a probability density function (p.d.f.) given by

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

wherever this derivative exists.

- If $f(y)$ is the p.d.f. of a continuous random variable Y, the c.d.f. of Y can be written as

$$F(y) = \int_{-\infty}^{y} f(x)dx$$

Properties of p.d.f.

- $f(x) \geq 0$ for all y.
- $\int_{-\infty}^{\infty} f(y)dy = 1$.

Calculating Probabilities from c.d.f. and p.d.f.

- If a random variable Y has a c.d.f. $F(y)$, then
  1. $P(y_1 < Y \leq y_2) = F(y_2) - F(y_1)$ for any $y_1 < y_2$,
  2. $P(Y = y) = F(y) - F(y-)$ for any y, where $F(y-)$ denotes the left limit of F at y.
- If Y has a p.d.f. $f(y)$, we can further write
  1. $P(y_1 < Y \leq y_2) = \int_{y_1}^{y_2} f(y)dy$ for any $y_1 < y_2$,
  2. $P(Y = y) = 0$ for any real y.

The Expected Value of a Random Variable

The mean or expected value of a random variable Y is denoted by $E(Y)$.

- Discrete case: Suppose Y is a discrete random variable with p.m.f. $p(y)$. Then $E(Y) = \sum_y yp(y)$. The $E(Y)$ is defined if $\sum_y |y|p(y)$ is finite.
- Continuous case: Suppose Y is a continuous random variable with p.d.f. $f(y)$. Then $E(Y) = \int_{-\infty}^{\infty} yf(y)dy$. The $E(Y)$ is defined if $\sum_y |y|f(y)dy$ is finite.

Expected Value of a Function of a Random Variable

If $g(Y)$ is a function of a random variable Y with p.m.f. $p(y)$ or with a p.d.f. $f(y)$, then the expected value of $g(Y)$ is given by

- Discrete case: $E(g(Y)) = \sum_y g(y)p(y)$
- Continuous case: $E(g(Y)) = \int_{-\infty}^{\infty} g(y)f(y)dy$.

Variance and SD of a Random Variable

- The variance of a random variable Y, denoted by $V(Y)$, is defined as $V(Y) = E(Y - \mu)^2$, where $\mu = E(Y)$.
- $SD(Y) = \sqrt{V(Y)}$.

Some Properties of Expected Value

1. If c is a constant, $E(c) = c$.
2. If $X_1, X_2, \ldots, X_k$ are random variables defined on the same sample space and $c_1, c_2, \ldots, c_k$ are constants, then $E\left(\sum_{i=1}^{k} c_i X_i\right) = \sum_{i=1}^{k} c_i E(X_i)$.
3. Special cases:

   (a) If $g_1(Y), g_2(Y), \ldots, g_k(Y)$ are functions of Y then $E\left(\sum_{i=1}^{k} g_i(Y)\right) = \sum_{i=1}^{k} E(g_i(Y))$. *Proof:* Take $X_i = g_i(Y)$ and $c_i = 1$ for $i = 1, 2, \ldots, k$. Then apply the previous result.

   (b) If a and b are two constants, then $E(a+bY) = a + bE(Y)$.

Some Properties of Variance

- If a and b are two constants, then $V(a + bY) = b^2 V(Y)$.
- $V(Y) = E(Y^2) - (E(Y))^2$.

Binomial Probability Model

- A fixed number (n) of trials.
- Trials are independent.
- Each trial has two possible outcomes – either a success or a failure.
- Probability of a success in a single trial is p.

Binomial Random Variable and Binomial Distribution

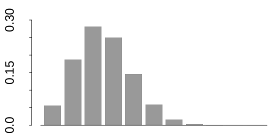
- In a binomial probability model let the random variable X denote the number of successes. Then X is a binomial random variable and the distribution of X is called a binomial distribution and denoted by Binomial (n,p). It has two parameters, n and p.
- The p.m.f. of a binomial random variable is given by

  $$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n,$$

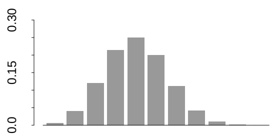
  where p and n are as described in the binomial model.

Proof: Let A be the event that the number of successes is x. Then number of sample points in A is $\binom{n}{x}$. Each sample point in A has probability $p^x(1-p)^{n-x}$.

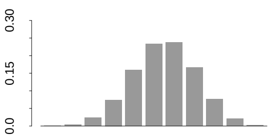
- **Some Examples**



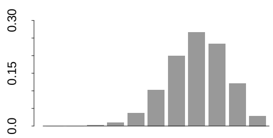
Binomial prob. hist. with n=10 and p=0.25



Binomial prob. hist. with n=10 and p=0.4



Binomial prob. hist. with n=10 and p=0.55



Binomial prob. hist. with n=10 and p=0.7

- The binomial table in the book gives the c.d.f. of the binomial distribution for different n and p.

- If X is a binomial random variable with parameters n and p, respectively, then $E(X) = np$ and $V(X) = np(1-p)$.

Examples

- <u>Example 1:</u> A box contains 1 green and 9 red balls. From that box 5 balls are drawn at random with replacement. What is the probability that exactly 2 draws will give red balls?

  Solution: In this experiment, getting a red ball in a draw is like a success in the binomial model. The draws are independent, the number of draws is fixed and in each draw the probability of getting a red ball is same $(\frac{9}{10})$. So, we have a binomial model. By Y denote the number of red balls in 5 draws. Then Y is a binomial random variable with $n = 5$ and $p = \frac{9}{10}$.

  - Calculation by using formula: $P(Y = 2) = \binom{5}{2}\left(\frac{9}{10}\right)^2\left(1-\frac{9}{10}\right)^{5-2} = 0.0081.$
  - Calculation by using table: From the binomial table, $P(Y \le 2) = 0.009$ and $P(Y \le 1) = 0$. Then $P(Y = 2) = P(Y \le 2) - P(Y \le 1) = 0.009$.

- <u>Example 2:</u> In the above experiment find the probability that al least 3 draws will give red balls.

  Solution: By formula, $P(Y \ge 3) = \sum_{y=3}^{5} P(Y = y) = 0.0729 + 0.32805 + 0.59049 = 0.99144.$

  By table: $P(Y \ge 3) = 1 - P(Y \le 2) = 1 - 0.009 = 0.991$.

Poisson Distribution and Poisson Random Variable

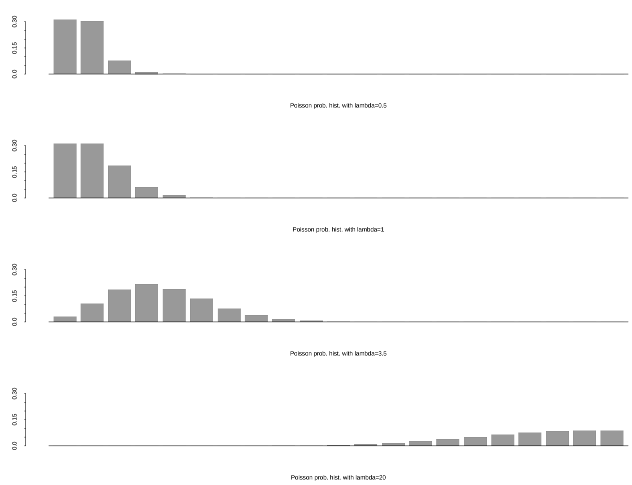
- Poisson probability distribution is a good model for the number of rare events that occur in a specific time or space.
- A Poisson distribution has a parameter λ which is the average rate of the events.
- Example: Number of accidents at a particular intersection during a month, number of errors made by a typist in typing a page, the number of cars entering a parking lot in a 5 minute period, *etc.*
- X is a Poisson random variable if it has a Poisson distribution and the p.m.f. of X, with average rate

parameter λ, is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots, \ \lambda > 0.$$

- It is easy to check that the above function is actually a p.m.f. because e^λ has a series expansion $\sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$.
- If X has a Poisson distribution with parameter λ, then $E(X) = \lambda$ and $V(X) = \lambda$.
- The Poisson table at the end on the book gives the c.d.f. of Poisson distribution for different λ.

Some Examples



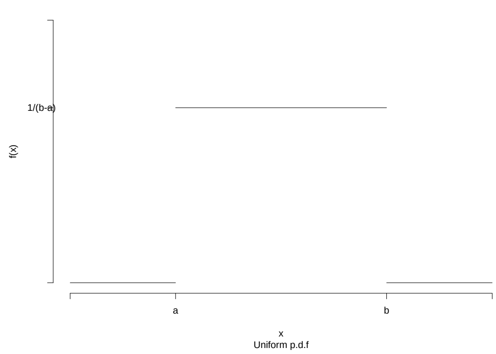
Poisson Approximation of Binomial Distribution

A Binomial (n,p) distribution for a large n and small p can be approximated by a Poisson distribution with parameter $\lambda = np$.

Uniform Distribution

- If X is distributed uniformly on the interval (a, b), then X has p.d.f.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$



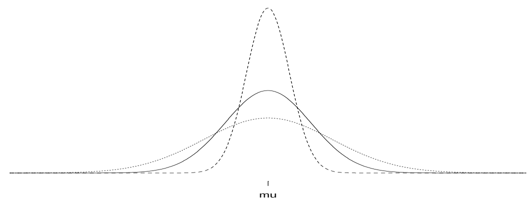
- If X is distributed uniformly on the interval (a, b), then

1. $E(X) = \frac{a+b}{2}$,
2. $V(X) = \frac{(b-a)^2}{12}$.

Example: Bill doesn't get up as soon as his alarm goes off. The extra time he sleeps in is given by the random variable X, which is distributed uniformly on the interval (0 min, 15 min).

- What's the probability that his extra sleeping time is less than 5 min?
- What's the probability that his extra sleeping time is less than 10 min, but more than 7 min?

Normal Distribution



- A normal distribution has two parameters mean μ and variance σ^2.
- A normal distribution with mean μ and variance σ^2 is denoted by $N(\mu, \sigma^2)$ and has p.d.f $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{\frac{-1}{2\sigma^2}(x-\mu)^2\}$, for $-\infty < x < \infty$.
- Has the classic bell-shape.
- Forms the basis of the empirical rule.
- Used to approximate many real-life variables.

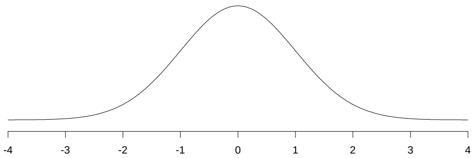
The Standard Normal Distribution

- A random variable Z has a standard normal distribution if it is distributed normally with $\mu = 0$ and $\sigma = 1$.
- Values of Z correspond to how many standard deviations away from the mean they are.
- Any normally distributed random variable can be transformed to the standard normal using this idea of "how many SD from the mean is it?"
- For $X \sim N(\mu, \sigma^2)$, we can transform X into standardized scores (z-scores) using $Z = \frac{X-\mu}{\sigma}$. Then $Z \sim N(0, 1)$.

Areas Under the Normal Curve

- To find $P(a \leq X \leq b)$, where $X \sim N(\mu, \sigma^2)$, we need to evaluate $\int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\{\frac{-1}{2\sigma^2}(x-\mu)^2\}dx$.
- But there is no closed-form solution to the integral!
- Numerical integration methods must be used in order to find the integral.
- We employ the fact that any normal distribution can be transformed into a standard normal distribution. Area under any normal curve can be found by first making the transformation to standard normal distribution and then finding the corresponding area under the standard normal curve.

Using z-scores and the Normal Table



Standard normal p.d.f

- Areas under the curve have been calculated and recorded in the normal table.
- For each z you look up in the table, you will get $P(Z \geq z)$
- Since the standard normal is symmetric about $\mu = 0$, no negative values for z are given on the table.
- Two important properties to remember when using the normal table: **1.** Symmetry and **2.** The total area under the curve is 1.

Using the Normal Table

Suppose the variable Z has a standard normal distribution, i.e., $Z \sim N(0, 1)$.

1. Find the following probabilities:
   - $P(-1 \leq Z \leq 1)$
   - $P(Z \leq -1.96)$
   - $P(-0.50 \leq Z \leq 1.25)$

2. Which value marks the 95th percentile?

3. Which values are the boundaries for the middle 80% of the data?

Example Using the Normal

The time required to complete a college achievement test was found to be normally distributed, with a mean of 110 minutes and a standard deviation of 20 minutes.

- What percentage of students will finish within 2 hours?
- What percentage of students will finish after 1.5 hours but before 2.5 hours?

- When should the test be terminated to allow just enough time for 90% of students to complete the test?

- What are the boundaries for IQR of the time it takes to complete the test?

Gamma Distribution

- If X has a gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, then X has p.d.f.:

$$f(x) = \begin{cases} \frac{x^{\alpha-1}e^{-\frac{y}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

where Γ is the gamma function defined on $(0, \infty)$ as

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy.$$

- Some Properties of gamma function:
  - $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for $\alpha > 0$.
  - $\Gamma(n) = (n-1)!$ for any integer $n \geq 1$.
  - $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

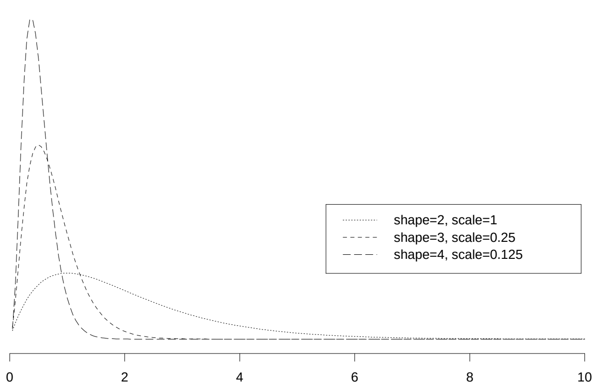
- Only when α is an integer, we can integrate the gamma p.d.f. over an interval and get a closed-form expression.

- For a gamma random variable X with parameters α and β,

  $E(X) = \alpha\beta, \quad V(X) = \alpha\beta^2.$

- Two special cases of the gamma have their own names.
  1. An exponential with parameter β is a gamma with $\alpha = 1$ and β.
  2. A chi-squared with ν degrees of freedom is a gamma with $\alpha = \frac{\nu}{2}$ and $\beta = 2$.

Various Gamma p.d.f.s



Beta Distribution

If X has a beta distribution with parameters $\alpha > 0$ and $\beta > 0$, then X has p.d.f.:

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$