Random Variables

Definition: A random variable X is a real valued function that maps a sample space S into the space of real numbers $\mathbb{R}$.

$$X : S \mapsto \mathbb{R}$$

A Random Variable is always " The number of ... "

As such, a random variable summarizes the outcome of an experiment in numerical form. For example, we may be interested in *how many* coin tosses resulted in heads rather than in the actual sequence of heads and tails from an experiment in which a coin is tossed many times.

There are two different kinds of random variables:

- DISCRETE random variables take on finitely many or countably infinitely many possible values.
- CONTINUOUS random variables take on uncountably infinitely many possible values, usually in an interval.

Definition: The probability mass function (PMF) of a discrete random variable is defined as the function

$$p(a) = P(X = a)$$

Since probability mass functions are probabilities, we must have $0 \leq p(a) \leq 1$ and if the probability mass function is summed over all possible values of X we must have

$$\sum_{\text{all } a} p(a) = 1$$

Probability mass functions are either specified in function form, e.g.,

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

or in table form, e.g.,

a	$p(a)$
0	1/4
1	1/2
2	1/4

Probability mass functions can also be graphed, most commonly as bar graphs with the possible values on the x-axis and bars whose heights represent the probabilities.

There is a handout with probability mass functions and other properties of specific discrete random variables that you are expected to be familiar with on the course website.

Example 36. Coupon collector problem

Suppose there are N distinct types of coupons and each time one obtains a coupon, it is, independently of the past, equally likely to be any one of the N types. One random variable of interest is T the number of coupons one has to collect in order obtain a complete set (with each type of coupon collected at least once). Find the probability mass function of T.

Definition: The function

$$F(a) = P(X \leq a)$$

is called the cumulative distribution function (CDF) of the random variable X. Cumulative distribution functions are always non-decreasing, right-continuous with

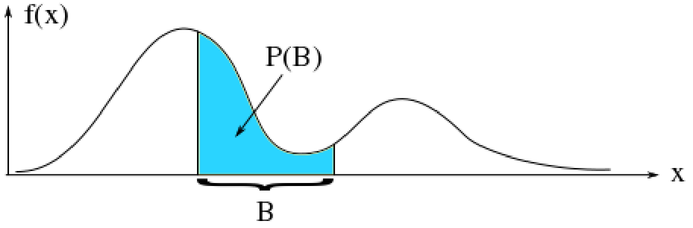
$$\lim_{b \to -\infty} F(b) = 0, \qquad \lim_{b \to \infty} F(b) = 1$$

Continuous Random Variables

Definition: Suppose X is a continuous random variable. Then a function $f(x)$ with

$$P(X \in B) = \int_B f(x)dx$$

is called the probability density function (PDF) of X.



Probability density functions must satisfy the following properties:

- They must be non-negative everywhere: $f(x) \geq 0$ for all $x \in \mathbb{R}$.
- If integrated over all possible values of X (or the whole of $\mathbb{R}$) we must obtain one.

$$\int_{-\infty}^{\infty} f(x) = P(X \in \mathbb{R}) = 1$$

since the random variable X must take on *some* value.

Example 37. Suppose X is a continuous random variable with PDF

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of the constant C.

(b) Find $P(X > 1)$.

Example 38. The cumulative distribution function of a random variable X is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \le x < 1 \\ \frac{2}{3} & 1 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

Graph the CDF. Is X discrete or continuous? Find

$P(X < 2)$

$P(X = 1)$

$P(X = 1/2)$

$P(1/2 \le X \le 1)$

Example 39. * An insurance company determines that N, the number of claims received in a week, is a random variable with

$$P(N = n) = \frac{1}{2^{n+1}}, \quad \text{where } n \ge 0$$

The company also determines that the number of claims in a given week is independent of the number of claims received in any other week. Calculate the probability that exactly seven claims will be received during a given two-week period.

Expected Value and Variance

Expected values were what gave rise to the study of probability originally. They can be understood as the "long-run-average" outcome or as the center of the distribution of a random variable. Expected values are defined very similarly for discrete and continuous distributions.

Definition: For a discrete random variable X with probability mass function $p(x)$, the expected value of X is defined as

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

That is, the expected value is a weighted sum of all possible values of X where the weights are the probabilities. You can think of expected values as the "center of gravity" of a distribution.

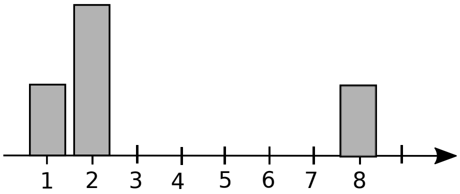
For a continuous random variable, the idea is very similar. But instead of taking a sum over the countably many possible values, the probability density function is integrated over all possible values.

Definition: Let X be a continuous random variable with probability density function $f(x)$. Then the expected value of X is defined as

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Example 40. Consider a discrete random variable with probability mass function $p(x)$. Find $E[X]$ and indicate it in the graph of the PMF.

x	$p(x)$
1	1/4
2	1/2
8	1/4



Note: An expected value must *always* be within the range of possible values of a random variable.

Example 41. * Find $E[X]$ if the density of the continuous random variable X is

$$f(x) = \begin{cases} \frac{|x|}{10} & -2 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Frequently, the expected value of a function of the random variable X is of interest, rather than the expected value of the random variable X itself.

Theorem: Let X be a discrete random variable and let $g(x)$ be a real valued function, then

$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$

Proof:

Example 42. Find $E[aX + b]$ in terms of $E[X]$ if $a, b \in \mathbb{R}$ are constants.

Example 43. * A tour operator has a bus that can accommodate 20 tourists. The operator knows that tourists may not show up, so he sells 21 tickets. The probability that an individual tourist will not show up is 0.02, independent of all other tourists. Each ticket costs 50, and is non-refundable if a tourist fails to show up. If a tourist shows up and a seat is not available, the tour operator has to pay 100 (ticket cost + 50 penalty) to the tourist. Calculate the expected revenue of the tour operator.

Similarly to the discrete case, we can also compute expected values of functions of continuous random variables. However, to prove the corresponding statement, the following result is helpful.

Lemma: For a nonnegative random variable Y (discrete or continuous)

$$E[Y] = \int_0^\infty P(Y > y)dy$$

Proof:

Theorem: If X is a continuous random variable with PDF $f(x)$, then for any real valued function $g(x)$,

$$E[g(X)] = \int_{-\infty}^\infty g(x)f(x)dx$$

Proof:

Example 44. A stick of length 1 is split at a point U having density function

$$f(u) = \begin{cases} 1 & 0 \le u \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Determine the expected length of the piece that contains the point p ($0 \le p \le 1$).

While the expected value of a random variable measures the "center" of the distribution of X, the variance measures the spread of the distribution.

Definition: If X is a random variable with mean μ, then the variance of X, denoted by $Var(X)$ is defined by

$$Var(X) = E[(X - \mu)^2]$$

That is, the variance is the expected squared difference of X and its mean.

Note: This definition is valid for both discrete and continuous random variables.

Fact: Alternatively, variance can be computed as

$$Var(X) = E[X^2] - E[X]^2$$

Proof: (discrete case)

Example 45. Let X be a continuous Uniform(a,b) random variable with density function

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

Find the variance of X.

Example 46. Let X be a random variable (discrete or continuous) with variance σ^2. Find $Var(aX + b)$, where $a, b \in \mathbb{R}$ are constants.

Example 47. * A recent study indicates that the annual cost of maintaining and repairing a car in a town in Ontario averages 200 with a variance of 260. A tax of 20% is introduced on all items associated with the maintenance and repair of cars (i.e., everything is made 20% more expensive). Calculate the variance of the annual cost of maintaining and repairing a car after the tax is introduced.

Distribution of a Function of a Random Variable

Suppose you know the distribution of a random variable X. How do you find the distribution of some function $Y = g(X)$ of X?

Example 48. Let X be a continuous random variable with CDF $F_X(x)$. Find the CDF of $Y = X^2$.

Theorem: Let X be a continuous random variable with probability density function $f_X(x)$. Suppose that $g(x)$ is a strictly monotonic (increasing or decreasing), differentiable (and thus continuous) function of x. Then the random variable Y defined by $Y = g(X)$ has a probability density function given by

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)]|\frac{d}{dy}g^{-1}(y)| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

Proof:

Example 49. Let X be a continuous nonnegative random variable with density function f and let $Y = X^n$. Find f_Y, the density function of Y.

Named Discrete Distributions

You should be familiar with the discrete and continuous distributions introduced in Math 161A. Being familiar with a distribution includes knowing its parameters, possible values, and the probability mass function or density as well as the cumulative distribution function of the distribution. In addition, you should know (and be able to derive where appropriate) formulas for the mean and variance of each distribution. Recall, that all distributions covered in 161A (and their key properties) are listed on the two handouts "Named Discrete Distributions" and "Named Continuous Distributions" available on Canvas. The same information is available inside the front and back cover of your textbook.

Each distribution is useful to model random variables for specific kinds of situations. We will *briefly* review the distributions and the situations for which they are intended.

BERNOULLI: $X \sim$ Bernoulli(p). X models whether or not a single trial will result in a success.

$$p(1) = p, \qquad p(0) = 1 - p$$

Here, p is the success probability.

Fact: $E[X] = p, Var(X) = p(1-p)$

BINOMIAL: $X \sim$ Binomial(n, p) X models the number of successes in n independent trials, each of which will result in a success with probability p.

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

Fact: $E[X] = np, Var(X) = np(1-p)$.

Fact: The sum of n independent Bernoulli random variables with the same parameter p has a Binomial(n, p) distribution.

Example 50. In a U.S. presidential election, the candidate who gains the maximum number of votes in a state is awarded the total number of electoral college votes allocated to that state. The number of electoral college votes is roughly proportional to the population of that state. That is, a state with population n has roughly nc electoral votes. In which states does your vote have more average power in a close election? Here, average power is defined as the expected number of electoral votes that your vote will affect. Let's assume that the total population of the state you are in is odd $n = 2k + 1$.

HYPERGEOMETRIC: $X \sim$ Hypergeometric(n, m, N). X is the number of special objects in a sample of size n taken from a population of N objects (without replacement) of which m are special.

$$p(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \ldots, \min\{n, m\}$$

Fact: $E[X] = n\frac{m}{N}, Var(X) = \left(\frac{N-n}{N-1}\right) n\frac{m}{N}\left(1 - \frac{m}{N}\right)$.

Fact: If N is very large compared to n, then it makes no difference whether you draw with or without replacement and thus a Hypergeometric random variable with $N >> n$ can be approximated by a Binomial random variable with $p = \frac{m}{N}$.

GEOMETRIC: $X \sim$ Geometric(p). X is the number of independent trials that have to be performed until the first success is observed. p is the probability of a success in each trial.

$$p(x) = (1-p)^{x-1}p, \qquad x = 1, 2, \ldots$$

Fact: $E[X] = \frac{1}{p}, Var(X) = \frac{1-p}{p^2}$.

Fact: The geometric distribution has the "lack-of-memory" property

$$P(X > s + t | X > t) = P(X > s)$$

Example 51. Find a closed-from formula for the CDF of a Geometric(p) random variable.

NEGATIVE BINOMIAL: $X \sim$ Negative Binomial(r, p). X is the number of independent trials that have to be performed until the r^{th} success is observed. p is the probability of success in each trial.

$$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad x = r, r+1, \ldots$$

Fact: $E[X] = \frac{r}{p}, Var(X) = r\frac{1-p}{p^2}$.

Fact: The sum of r independent geometric random variables (with the same parameter p) has a negative binomial distribution with parameters r and p.

Example 52. The Banach match problem

A pipe smoking mathematician carries two matchboxes, one in his right pocket and one in his left. Each time he needs a match, he is equally likely to choose either pocket. Initially, each box contained N matches. Consider the moment the mathematician first discovers a matchbox to be empty. At this time, what is the probability that there are exactly k matches in the other box ($k = 0, 1, \ldots, N$).

POISSON: $X \sim$ Poisson(λ). X is the number of times a "rare" event occurs (in a certain time or space interval).

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, \ldots$$

The Poisson distribution can be understood as an approximation of the binomial distribution in those cases where n is large and p is small enough so that np is moderately small.

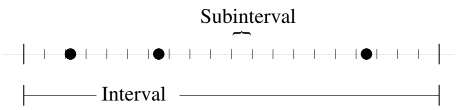
Examples:

- The number of hurricanes in the central U.S. in a month.
- The number of misprints on a page of some document.
- The number of walnuts in a walnut cookie.

Fact: $E[X] = \lambda, Var(X) = \lambda$.

Fact: The sum of independent Poisson random variables is Poisson. The parameters add.

Example 53. Poisson variables can be derived from Binomial random variables.To count the number of rare events, imagine an interval (that can stand for time, or a page, or a volume of cookie dough) split up into n little subintervals.



It is always possible to make the subintervals small enough (by making n large), such that there is at most one event in a subinterval. Suppose that the probability that a subinterval has an event in it is p.

We are interested in the number of times X the event occurs in the interval. Strictly speaking, X has a Binomial(n, p) distribution but with a very large n and a small p (since the events are "rare"). What happens to the Binomial PMF, if $n \to \infty$? Let $\lambda = np$.

While working with specified distributions is quite straightforward (e.g., computing values of the PMF, CDF, expected values or variances) it can sometimes be challenging for students to recognize which distribution to use in a specific situation.

Example 54. Skittles are small fruit candy that come in many different colors. About 10% of all skittles are orange. Skittles are sold in randomly filled packages of 90 candy each. For each of the following random variables, state the distribution and find the values of all relevant parameters.

(a) X is the number of orange skittles in one package.

(b) X is the number of packages you buy until you get one that has no orange skittles.

(c) X is the number of orange skittles you eat, if you eat ten from a full package that had 12 orange ones in it.

(d) Suppose you get "Skittle-cravings" on average twice every day. X is the number of skittle-cravings you'll have within the next 36 hours.

(e) X is the number of skittles you'll eat (randomly selected from a very large supply) until you eat your 10^{th} orange skittle.

(f) Suppose you eat ten randomly selected skittles every day for a week. X is the number of days on which you eat no orange ones.

Named Continuous Distributions

UNIFORM: $X \sim \text{Uniform}(a, b)$. The random variable X is equally likely to assume any position in the interval $[a, b]$.

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

Fact: $E[X] = \frac{b+a}{2}, Var(X) = \frac{(b-a)^2}{12}$.

Fact: The CDF of a continuous uniform(a, b) random variable is

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

EXPONENTIAL: $X \sim \text{Exponential}(\lambda)$. The exponential distribution is the continuous analog to the geometric distribution. It is frequently used to model waiting times and has a close relationship with the Poisson distribution.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Examples:

- X is the time until the next customer arrives at a bank.
- X is the time until a lightbulb burns out (lifetime).
- X is the mileage you get out of one tank of gas.

Fact: $E[X] = \frac{1}{\lambda}, Var(X) = \frac{1}{\lambda^2}$.

Fact: Like the geometric distribution, the exponential distribution also has the memoryless property

$$P(X > s + t | X > t) = P(X > s)$$

Example 55. Suppose the number of events that occur in a unit time interval has Poisson distribution with mean λ. Find the distribution of the amount of time until the first event occurs.

GAMMA DISTRIBUTION: $X \sim \text{Gamma}(r, \lambda)$. X models the continuous waiting time until the r^{th} occurrence of an event.

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{r-1}}{\Gamma(r)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Here, $\Gamma(r)$ is the gamma function which is defined as

$$\Gamma(r) = (r-1)!$$

if r is an integer. But r does not necessarily have to be an integer in the above definition of the gamma distribution. For general r, the gamma function is defined as

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

Fact: $E[X] = \frac{r}{\lambda}, Var(X) = r\frac{1}{\lambda^2}$.

Fact: The sum of r independent exponential random variables with the same parameter λ has a gamma(r, λ) distribution.

Fact: A gamma random variable with $\lambda = \frac{1}{2}$ and $r = \frac{n}{2}$ for some positive integer n is called a χ^2_n (chi-squared) random variable with n degrees of freedom.

NORMAL: $X \sim \text{Normal}(\mu, \sigma^2)$. The normal distribution is also sometimes called the Gaussian distribution after Carl Friedrich Gauss who was the first to officially define this distribution in 1809. Its PDF has the characteristic "bell-curve-shape".

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad -\infty < x < \infty$$

Fact: $E[X] = \mu, Var(X) = \sigma^2$.

Example 56. Linear transformations of normal random variables are normal. That is, suppose $X \sim \text{Normal}(\mu, \sigma^2)$. Find the distribution of $Y = aX + b \quad (a, b, \in \mathbb{R})$.

Fact: A normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called a standard normal distribution.

Fact: The CDF of a standard normal random variable is denoted $\Phi(x) = P(X \leq x)$ where $X \sim$ Normal(0,1). There is no closed form for $\Phi(x)$. Instead, the values of $\Phi(x)$ are obtained from tables or through software.

Fact: The sum of independent normal random variables is normal. Let $X \sim$ Normal(μ_1, σ_1^2) and $Y \sim$ Normal(μ_2, σ_2^2) be independent, then

$$X + Y \sim \text{ Normal}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

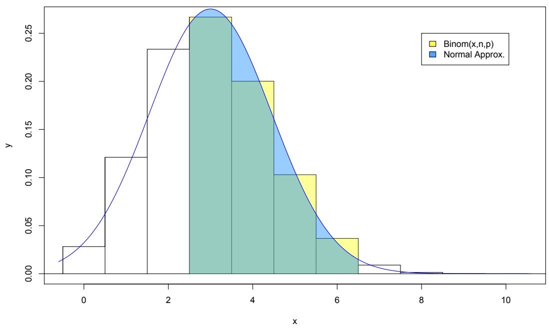
Even before Carl Friedrich Gauss officially defined the normal distribution (in his paper about least squares and maximum likelihood methods), Abraham deMoivre and Pierre-Simon Laplace showed that a Binomial random variable with large n can be approximated well by a normal random variable with the same mean and variance as the binomial. They first proved this result only for $p = \frac{1}{2}$ and after Gauss' paper extended it to the general case.

Theorem: The DeMoivre-Laplace Limit Theorem
Let S_n denote the number of successes in n independent trials, each resulting in a success with probability p, then for any $a < b$

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \xrightarrow[n\to\infty]{} \Phi(b) - \Phi(a)$$

We will consider the proof when we discuss the Central Limit Theorem (of which this is a special case) later in the semester.

Example 57. Consider $X \sim$ Binomial$(n = 10, p = 0.3)$. Use your calculator to compute $P(3 \leq X \leq 6)$. Also use the normal approximation to the Binomial (with continuity correction) to approximate this same probability. Use either your calculator or a normal table to look up the required normal CDF values.



Chi-Squared Distribution

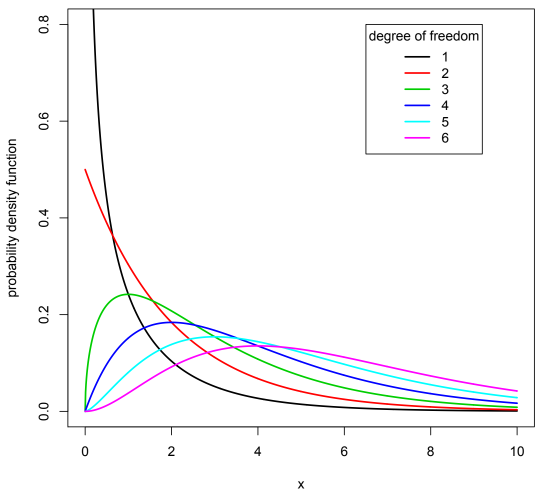
The χ^2 (chi-squared) distribution appears in several common hypothesis tests (e.g., t-test, goodness of fit, likelihood ratio). It is related to both the gamma and the normal distributions.

Definition: A continuous random variable with density function

$$f(x) = \begin{cases} \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

is said to have a χ^2_n distribution (chi-squared with n degrees of freedom).

Fact: The χ^2_n-distribution is a special case of the gamma distribution with $\lambda = \frac{1}{2}$ and $r = \frac{n}{2}$.



Fact: The mean and variance of a χ^2_n distribution are

$E[X] =$ $\qquad\qquad$ $Var(X) =$

Example 58. Let $Z \sim$ Normal(0,1). Show that Z^2 has a χ^2_1 distribution.

Fact: The sum of independent χ^2 random variables has a χ^2 distribution and the degrees of freedom add. That is, let $X_1 \sim \chi^2(df = n_1)$ and $X_2 \sim \chi^2(df = n_2)$ be independent, then

$$X_1 + X_2 \sim \chi^2(df = n_1 + n_2)$$

Note: This will become easier to prove, once we learn about Moment Generating Functions in Chapter 7.

Example 59. Suppose $Z_1, Z_2, \ldots, Z_n$ are independent and identically distributed standard Normal random variables, then what is the distribution of $Z_1^2 + Z_2^2 + \cdots + Z_n^2$?