Random Variables (general facts)

Note: All sums and integrals are taken over the set of all possible values of the random variable X unless stated otherwise. In the discrete case, X has probability distribution function $p(X = x)$ or $p(x)$ for short. In the continuous case, X has probability density $f(x)$.

- Computing probabilites of events

  Discrete case: $P(E)$ =the sum of the probabilities of all of the values of X in the event E

  Continuous case: $P(a < X < b) = \int_a^b f(x)dx$ even if one or both of a and b are infinite

- **Expected Value** = **Mean** of a random variable (definition)

  Discrete case: $E(X) = \mu_X = \sum xp(x)$

  Continuous case: $E(X) = \mu_X = \int xf(x)dx$

- **Variance** (definition)

  Discrete case: $\sigma_X^2 = \sum (x - \mu_X)^2 p(x)$

  Continuous case: $\sigma_X^2 = \int (x - \mu_X)^2 f(x)dx$

- **Variance** (computational formula; this is a Theorem)

  In general: $\sigma_X^2 = E(X^2) - (\mu_x)^2$

  Discrete case: $\sigma_X^2 = \left[\sum (x^2)p(x)\right] - (\mu_x)^2$

  Continuous case: $\sigma_X^2 = \left[\int (x^2)f(x)dx\right] - (\mu_x)^2$

- **Standard Deviation** $\sigma = \sqrt{\sigma^2}$ in both cases (definition)

- These formulas give the definitions of the mean, variance, and standard deviation for a random variable. In many cases, the special distributions described below have simpler formulas for their means and standard deviations, which are proved using these definitions. The formulas are given below as well.

Discrete Uniform Distribution

"Uniformly distributed" means all possible values of X are equally likely.

- If there are n possible values of X, the probability of each one is $\frac{1}{n}$

- If the possible values of X are the integers 1 through n, more can be said:

  $\mu_X = \frac{n+1}{2}$

  $\sigma_X^2 = \frac{n^2-1}{12}$

  Note that these formulas hold **only** when the possible values of X are the integers $\{1, 2, ..., n\}$: they do not hold in general.

Binomial Distributions

Situation: We repeat a simple experiment or choice called a Bernoulli trial (or just "trial"), which has only two possible outcomes, called "success" and "failure". The probability of "success" is the same on each repetition of the trial. We repeat the trial a certain number of times, and want to know the probability of getting a certain number of "successes".

Two other situations related to this can also be modeled by the binomial distribution:

- Sampling with replacement
- Sampling from an infinite population

The binomial distributions are a 2-parameter family of discrete distributions with parameters n= the number of repetitions of the trial, and p= the probability of success on a single trial; the possible values of X are the integers 0 through n.
This describes the distribution of the number of successes for an experiment which consists of n repeated independent trials, each having two possible outcomes (success and failure). The probability of success on one trial is p (sometimes called π).

- the probability distribution function is $b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$
- $\mu_X = np$
- $\sigma_X^2 = np(1-p)$
- Define a random variable $Y = \frac{X}{n}$ representing the fraction of the time one saw successes; then

  $\mu_Y = p$
  $\sigma_Y^2 = \frac{p(1-p)}{n}$
  This random variable is used in proving theorems concerning the proportion of successes for a binomial random variable, for example, the Law of Large Numbers.

Hypergeometric Distribution

Situation: We are taking a random sample (without replacement) from a finite population. There is some attribute that we are interested in, called "success". We want to know the probability of getting a certain number of "successes" in the sample.

The Hypergeometric distributions are a 3-parameter family of discrete distributions with parameters N= the size of the population from which the sample is taken, M= the number of successes in the population, and n=the size of the sample. X represents the number of successes in a sample of size n chosen at random and without replacement. The possible values of X are the integers $0, 1, 2, ..., n$.

- the probability distribution function is

$$h(x; N, M, n) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

- $\mu_X = n(\frac{M}{N})$
- $\sigma_X^2 = n(\frac{M}{N})(\frac{N-M}{N})(\frac{N-n}{N-1})$ [I group the factors this way for a reason: see next item.]
- Note that if we set $p = \frac{M}{N}$ then these are nearly the same as for the binomial distribution:

$$\frac{N-M}{N} = 1 - \frac{M}{N} = 1 - p$$

  . This shows that the hypergeometric distribution can be estimated by the binomial distribution when N is very large compared to n, as the remaining factor $\frac{N-n}{N-1}$ is close to 1 in that case.

  Usual rule of thumb: estimate by binomial when n is not more than 5% of N.

  Estimate by $b(x; n, p)$ with $p = \frac{M}{N}$

Poisson Distribution

Situation: We have a random process, called a Poisson process, which satisfies the following conditions:

- The outcome of the process is the number of “successes” that occur in a given interval of time (which may not always be literally time: see below)
- In any very small interval of time, the probability of getting more than one “success” is essentially 0.
- The numbers of“successes” in two mutually exclusive intervals of time are independent.
- The probability of “success” in any interval of time is proportional to the length of the interval.

The Poisson distributions are a 1-parameter family of discrete distributions with parameter λ= the average number of success in a certain “unit of time” (which may not be literally time, but may be an length, area, volume, or a fixed number of repetitions of a binomial trial, for example). X represents the number of successes in that fixed unit of “time”. The possible values of X are the whole numbers.

The Poisson distributions can also be thought of as a way of approximating the binomial distributions when n is large and p is small in such a way that $np = \lambda$

- the probability distribution function is

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

- $\mu_X = \lambda$
- $\sigma_X^2 = \lambda$
- We can sometimes approximate the binomial distribution $b(x; n, p)$ by the Poisson distribution:

  A common rule of thumb: $n \geq 20$ and $p \leq 0.05$

  Take $\lambda = np$

- The waiting time for the first success (or the waiting time between successes) for the Poisson distribution has an exponential distribution with $\theta = \frac{1}{\lambda}$. See below.

Negative Binomial Distribution or Waiting Time Distribution or Pascal Distribution

Situation: the same as for the binomial distribution, except that we are interested in the probability that it takes a certain number of repetitions of the trial before we see a certain number of successes. This is called the "waiting time for the k-th success". (We think of the number of repetitions as a measure of time.)

When $k = 1$ we call this the **Geometric distribution** $g(x,p)$
This gives the probability that it takes x trials until the first "success" occurs.

$g(x,p) = p(1-p)^{x-1}$

$\mu_X = \frac{1}{p}$ and $\sigma_X^2 = \frac{1}{p}\left(\frac{1}{p} - 1\right) = \frac{1}{p^2} - \frac{1}{p}$ for the geometric distribution

The negative binomial distributions are a 2-parameter family of discrete distributions with parameters p= the probability of success on a single trial and k= the total number of successes; X represents the number of trials needed until we see k successes (in other words, the "waiting time" until k successes, although it may not be literally time). The possible values of X are the integers $k, k+1, k+2, ...$

- the probability distribution function is $b^*(x;k,\theta) = \binom{x-1}{k-1}p^k(1-p)^{x-k}$
- $\mu_X = \frac{k}{p}$
- $\sigma_X^2 = \frac{k}{p}(\frac{1}{p} - 1)$

Two useful theorems

- $b(x;n,p) = b(n-x;n,1-p)$ (interchanges the role of "success" and "failure")
- $b^*(x;k,p) = \frac{k}{x}b(k;x,p)$ (computes b^* by using b)

Continuous Uniform Density

The continuous uniform densities are defined on an interval (which must be a finite interval).

- If the interval is (α, β) or $[\alpha, \beta]$ then the density function is $f(x) = \frac{1}{\beta-\alpha}$ for $\alpha < x < \beta$ (The density function is a constant function whose value is [1 over the length of the interval].)
- $\mu_X = \frac{\alpha+\beta}{2}$, the average of α and β (which is the midpoint of the interval)
- $\sigma_X^2 = \frac{1}{12}(\beta - \alpha)^2$
- Probabilities are easy to compute: you do not need to integrate (since the density is a constant function). For example, the probability of an interval (a, b) or $[a, b]$which is a subinterval of (α, β) is just the length of the interval times $f(x) = \frac{1}{\beta-\alpha}$: $P(a < x < b) = \frac{b-a}{\beta-\alpha}$ as long as $\alpha \le a \le b \le \beta$.

Gamma Distribution

The Gamma distributions are a 2-parameter family of continuous densities with parameters α and β. They are not given a specific interpretation in general, but some special cases (two of which are described below) are important in practical applications.

- the probability density function is

$$g(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \text{ for } x > 0$$

- $\mu_X = \alpha\beta$
- $\sigma_X^2 = \alpha\beta^2$
- $\Gamma(x)$ is the Gamma function, which generalizes the factorial:

  $\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy$ for $x > 0$ [definition of the Gamma function]

  $\Gamma(x)$ satisfies the recursion relation $\Gamma(x) = (x - 1)\Gamma(x - 1)$

  $\Gamma(1) = 1$ as you can easily compute

  So $\Gamma(n) = (n - 1)!$ for any natural number n

- If α is a positive integer, the Gamma densities can be integrated by using integration by parts as many times as necessary. If α is not a positive integer, it is not usually possible to find the integrals in closed form: numerical methods must be used to estimate the integral in that case.

Exponential Distribution

The exponential distributions are a one-parameter family of densities with parameter θ. They are Gamma distributions with $\alpha = 1$ and $\beta = \theta$.

- the probability density function is
$$f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}} \text{ for } x > 0$$
- $\mu_X = \theta$
- $\sigma_X^2 = \theta^2$
- Also note that these densities are easy to integrate: If a and b are positive real numbers, as you can easily check we have:
$$\int_a^b \frac{1}{\theta}e^{-\frac{x}{\theta}}\,dx = e^{-\frac{a}{\theta}} - e^{-\frac{b}{\theta}}$$
$$\int_0^b \frac{1}{\theta}e^{-\frac{x}{\theta}}\,dx = 1 - e^{-\frac{b}{\theta}}$$
$$\int_a^\infty \frac{1}{\theta}e^{-\frac{x}{\theta}}\,dx = e^{-\frac{a}{\theta}}$$
- One use of the exponential distributions is to describe the waiting times for a Poisson process:

The waiting times for the first success, or the waiting time between successes, of a Poisson process with mean λ, has exponential probability density with $\theta = \frac{1}{\lambda}$

In other words, the waiting time density is the exponential density $f(x; \frac{1}{\lambda}) = \lambda e^{-\lambda x}$; its mean is $\mu_X = \frac{1}{\lambda}$ and its variance is $\sigma_X^2 = \frac{1}{\lambda^2}$

To compute the probability of having to wait k units of time before the first success (or between successes) we integrate this density from 0 to k.

Chi-Square Distribution

The Chi-Square distributions (χ^2 distributions) are a one-parameter family of densities with parameter ν. They are Gamma distributions with $\alpha = \frac{\nu}{2}$ and $\beta = 2$.

- the probability density function is
$$f(x) = \frac{1}{2^{\frac{\nu}{2}}}x^{\frac{\nu-2}{2}}e^{-\frac{x}{2}} \text{ for } x > 0$$
- $\mu_X = \nu$
- $\sigma_X^2 = 2\nu$
- The parameter ν is called the number of degrees of freedom of the distribution.
- This distribution has a number of applications in sampling theory; for example, it describes the distribution of the sample variances of a random variable which has a normal distribution.