

Probability Notes

Eric Desoiza

2.1 Sample Space and Events

A specific subset of a sample space is an event and is a set consisting of possible outcomes of the experiment

$A = \cup_{i=1}^{\infty} A_i$ is a demonstration of all the specific outcomes in the subsets

A^c is considered the complement of A, which is all outcomes in this sample space not within A

The complement of the sample space is an empty set as all possible outcomes are located within the sample space and thus we have that null set. If AB is a null set then it means they cannot both occur at once and are mutually exclusive

2.2 Axioms of Probability

There are three primary axioms when it comes to the probability of an event in an experiment:

Axiom 1: $0 \leq P(A) \leq 1$

Axiom 2: $P(S) = 1$

Axiom 3: For an subsequence of mutually exclusive events:

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i), \quad n = 1, 2, \dots, \infty$$

The last axiom demonstrates that with their mutual exclusivity the probability of the union of events is the same as the sum of events

These axioms also lead to the understanding that $P(A^c) = 1 - P(A)$

2.3 Conditional Probability and Independence

In conditional probability, what is conditioned for becomes the sample space and the probability of the proposed event is based on that sample space.

$$P(A|B) = \frac{P(AB)}{P(B)}$$

It is important to note that $A = AB \cup AB^c$

Because AB and AB^c are mutually exclusive, the preceding yields

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

When we utilize the preceding formula, we sat that we are computing $P(A)$ by condition on whether or not B occurs

We can also comptue the prbability of an event A by conditioning on which of the B_i occur. The formula is:

$$A = AS = A(\cup_{i=1}^n B_i) = \cup_{i=1}^n AB_i$$

which implies:

$$\sum_{i=1}^n P(A|B_i)P(B_i)$$

2.4 Random Variables

The quantities of interest that are determined by the results of the experiment are known as random variables.

The cumulative distribution function for a random variable X for any value x is:

$$F(x) = P\{X \leq x\}$$

A random variable can be fintie or a countable number of values is said to be discrete. we define the discrete random variable X with $p(x)$ by:

$$p(x) = P\{X = x\}$$

and because of these discrete amount of values we come to

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

In a continuous case in which the possible value lies in a interval or set of numbers we define it as:

$$P\{X \in C\} = \int_C f(x)dx$$

The cumulative distribution is done by integration and reversal can be done by differentiation

This leads us to continuous joint distribution functions which are cumulatively defined as:

$$F(x, y) = P\{X \leq x, Y \leq y\}$$

Thus specifying the probability of X being less than or equal to a value and Y less than or equal to a value.

In a discrete case we have the following

$$p(x, y) = P\{X = x, Y = y\}$$

Whereas in a continuous case:

$$P\{X \in C, Y \in D\} = \int_D \int_C f(x, y)dx dy$$

These random variables are independent if for any set of real numbers C and D

$$P\{X \in C, Y \in D\} = P\{X \in C\}P\{Y \in D\}$$

Loosely speaking, X and Y are independent if knowing the value of one of them does not affect the probability distribution of the other.

For all x,y

$$f(x, y) = f_X(x)f_Y(y)$$

where $f_X(x)$ and $f_Y(y)$ are the density functions of X and Y, respectively.

2.5 Expectation

$$E[X] = \sum_i x_i \{P = x_i\}$$

For a continuous random variable:

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

For $E[g(X)]$ it seems that it should be a weighted average of the possible values of g(x) much like a composition of functions and thus the value of x used in calculating expected value

$$E[g(X)] = \sum_x g(x)p(x)$$

For a continuous random variable:

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

Corollary: If a and b are constants then

$$E[aX + b] = aE[X] + b$$

Additionally we can see that expectation is a linear operation in the cases of the addition of two random variables

2.6 Variance

Definition: if X is a random variable with mean μ , then the variance of X, denoted by $\text{Var}(X)$, is defined by

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

That is

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

additionally we can know that

$$Var(aX + b) = a^2 Var(X)$$

This shows that Variance is not a linear operation and does not abide by the same process as expectation.

Definition: The covariance of two random variables X and Y, denoted $Cov(X,Y)$ is defined by

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

where $\mu_x = E[X]$ and $\mu_y = E[Y]$

Using the linearity of expectations we arrive to the following equation

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

We can also derive an additional formula for the Variance of the sum of two random variables

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

An important note on Covariance is that if two variables are independent their variance is Zero. This is not a bijective statement as Zero covariance does not guarantee independence

Since we know that two independent variables will have zero covariance we can now deduce

$$Var(X + Y) = Var(X) + Var(Y)$$

And with independent variables being intersected, we can evaluate to a simpler understanding

$$E[XY] = E[X]E[Y]$$

The correlation between two random variables X and Y, denoted as $Corr(X,Y)$ is defined by

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

2.7 Chebyshev's Inequality and the Laws of Large Numbers

Markov's Inequality: if X takes on only nonnegative values, then for any value $a > 0$

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

We also have Chebyshev's inequality, which tells us that the probability of us finding a value differing from the mean more than k times its standard deviation is bound by $1/k^2$

Chebyshev's Inequality: If X is a random variable having mean μ and variance σ^2 , then for any value $k > 0$

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

We can now use Chebyshev's to prove the weak law of large numbers. This law states that the probability of the average of n many iid random variables differs from its mean by more than ϵ goes to zero as we approach infinity

The Weak Law of Large Numbers: Let X_1, X_2, \dots be a sequence of iid random variables having mean μ . Then for any $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \left(P \left\{ \left| \frac{X_1 + \dots + X_n}{n} \right| > \epsilon \right\} \right) = 0$$

A generalization of the weak law is the strong law of large numbers, which states that, with probability 1,

$$\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu$$

That is with certainty, the long-run average of iid random variables will converge to the mean.

2.8 Discrete Random Variables

Binomial Random Variables: Suppose that n independent trials, each of which results in a “success” with probability p , are to be performed. We provide this Random Variable with parameters (n, p) . Its probability mass function is given by

$$p_i \equiv P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \dots, n$$

where

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

A binomial $(1, p)$ random variable is called a Bernoulli.

We can also derive the expected value and variance of such a random variable

$$E[X_i] = p, \quad \text{Var}(X_i) = p(1-p)$$

And from this we can establish a binomial (n, p) random variable X ,

$$E[X] = np, \quad \text{Var}(X) = np(1-p)$$

This is since X_i should be independent

Recursive formula

$$p_{i+1} = \frac{n-i}{i+1} \frac{p}{1-p} p_i$$

Poisson Random Variables: A random variable that takes on one of the values $0, 1, 2, \dots$ is said to be a Poisson random variable with parameter $\lambda, \lambda > 0$, if its probability mass function is given by

$$p_i = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 0, 1, \dots$$

$$e = \lim_{n \rightarrow \infty} (1 + 1/n)^n$$

An usual example for such a random variable is to approximate the distribution of the number of success in a large number of trials (which should be independent or at most weakly dependent) where each trial has a small probability of success.

This means that with a binomial random variable we can apply a large n and small p and arrive to the approximate formula above for Poisson

$$E[X] = \text{Var}(X) = \lambda$$

The recursive formula would be that of

$$p_{i+1} = \frac{\lambda}{i+1} p_i, \quad i \geq 0$$

If there is a type one event and type two event in a Poisson Sample Space then we have

$$E[N_1] = \lambda p \quad E[N_2] = \lambda(1 - p)$$

Geometric Random Variables: Consider independent trials, each of which is a success with probability p . If X represents the number of the first trial that is a success, then

$$P\{X = n\} = p(1 - p)^{n-1}, \quad n \geq 1$$

this can be derived quite easilt since the first $n - 1$ trials must be a failure and result in the n th being a success. This is additionally true because of the independence of trials

The expected value and variance of such a random variable is as follows

$$E[X] = \frac{1}{p}, \quad \text{Var}(X) = \frac{1 - p}{p^2}$$

Negative Binomial Random Variable: If we let X denoted the number of trials need to amass a total of r successes when each trial is independently a success with probability p , then X is a negative binomial, or Pascal. The probability mass function is as follows

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1 - p)^{n-r}, \quad n \geq r$$

The first $n - 1$ trials must result in exactly $r - 1$ succeeded and then the n th trial must be a success

The expected value and variance of such a random variable is as follows

$$E[X] = \sum_{i=1}^r E[X_i] = \frac{r}{p}, \quad \text{Var}(x) = \frac{r(1 - p)}{p^2}$$

Hypergeometric Random Variables: Consider an urn containing $N + M$ balls, of which N are light colored and M are dark colored. If a sample size n is randomly chosen then X , the number of light colored balls selected, has probability mass function

$$P\{X = i\} = \frac{\binom{N}{i} \binom{M}{n-i}}{\binom{N+M}{n}}$$

The expected value and variance of such a random variable is as follows

$$E[X] = \frac{nN}{N+M}, \quad \text{Var}(X) = \frac{nNM}{(N+M)^2} \left(1 - \frac{n-1}{N+M-1}\right)$$

2.9 Continuous Random Variables

Uniform Distributed Random Variables: A random variable X is said to be uniformly distributed over interval (a,b) , $a < b$, if its probability density function is given by

$$f(x) = \frac{1}{b-a}, \quad \text{if } a < x < b$$

The first and second moment is as follows

$$E[X] = \frac{b+a}{2}$$

$$E[X^2] = \frac{a^2 + b^2 + ab}{3}$$

And the Variance is as such:

$$\text{Var}(X) = \frac{1}{12}(b-a)^2$$

The cumulative distribution is that of

$$F(x) = P\{X \leq x\} = \frac{x-a}{b-a}$$

Normal Random Variables: A random variable is said to be normally distributed with mean μ and variance σ^2 if its probability density function is given by

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

The expected values and variance of such random variables are as follows:

$$E[X] = \mu \quad \text{and} \quad \text{Var}(X) = \sigma^2$$

with an constants a and b , $aX + b$ is normally distributed with mean $a\mu + b$ and variance $a^2\sigma^2$.

$$Z = \frac{X - \mu}{\sigma}$$

is normal with mean 0 and variance 1. Such a random variable Z is a standard normal distribution. Let Φ denote the distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx, \quad -\infty < x < \infty$$

$$F(x) = P\{X \leq x\} = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

The wide applicability of normal random variables results from one of the most important theorems of probability theory. The Central Limit Theorem asserts the sum of a large of independent random variables has approximately normal distributions

Central Limit Theorem: Let X_1, X_2, \dots be a sequence of iid random variables having finite mean μ and finite variance σ^2

$$\lim_{n \rightarrow \infty} \left(P \left\{ \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} < x \right\} \right) = \Phi(x)$$

Exponential Random Variables: A continuous random variable having PDF

$$f(x) = \lambda e^{-\lambda x}, \quad 0 < x < \infty$$

Its CDF is given by

$$F(x) = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}, \quad 0 < x < \infty$$

The expected value and variance of such a random variable is as follows

$$E[X] = \frac{1}{\lambda} \quad \text{and} \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

Memoryless: The key property of the exponential random variable is that they possess a memoryless property, where we say that the nonnegative random variable X is memoryless if

$$P\{X > s + t | X > s\} = P\{X > t\} \text{ for all } s, t \geq 0$$

For example this means that the probability of the lifetime of some unit to last more than t would be the same if it has already lasted s time or not.

We also know that

$$P\{X > s + t\} = P\{X > s\}P\{X > t\}$$

It is also important to note that cX is exponential with parameter λ/c

Also $\min(X_1, \dots, X_n)$ is exponential with rate $\sum_i \lambda_i$ as is independent of which one of the X_i is the smallest.

Poisson Process and Gamma Random Variables: Suppose that events are occurring at random points in time and let N(t) denote the number of events that occur in the time interval [0,t]. These events are said to constitute a Poisson process having rate $\lambda, \lambda > 0$ if

$$\text{a) } N(0) = 0$$

- b) The numbers of events occurring in disjoint time intervals are independent
- c) The distribution of the number of events that occur in a given interval depends only on the length of the interval and not on its location
- d) $\lim_{h \rightarrow 0} \frac{P\{N(h)=1\}}{h} = \lambda$
- e) $\lim_{h \rightarrow 0} \frac{P\{N(h) \geq 2\}}{h} = 0$

Condition a states that the process begins at time 0. Condition b, the independent increment assumption, states that the number of events by time t is independent of the number of events that occur between t and $t+s$. Condition c, the stationary increment assumption, states that the probability distribution of $N(t+s) - N(t)$ is the same for all t . Conditions d and e state that in a smaller interval of length h , the probability of one event occurring is λh , where as the probability of two or more is approximately 0.

The interarrival times X_1, X_2, \dots are independent and identically distributed exponential random variables with parameter λ

Definition: A random variable having PDF

$$f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}, \quad x > 0$$

is said to be a gamma random variable with parameters (n, λ) .

The sum of n independent exponential random variables, each having parameter λ , is a gamma function with parameters (n, λ) .

Nonhomogeneous Poisson Process: The major weakness of the Poisson process is its assumption of equally likely events in all intervals of equal size. Something that relaxes this assumption is that of a nonhomogeneous or nonstationary process

We say that $\{N(t), t \geq 0\}$ constitutes a nonhomogeneous Poisson process with intensity factor $\lambda(t), t \geq 0$ if

- a) $N(0) = 0$
- b) The number of events that occur in disjoint time intervals are independent
- c) $\lim_{h \rightarrow 0} P\{\text{exactly 1 event between } t \text{ and } t+h\}/h = \lambda(t)$
- d) $\lim_{h \rightarrow 0} P\{2 \text{ or more events between } t \text{ and } t+h\}/h = 0$

The function $m(t)$ is defined by

$$m(t) = \int_0^t \lambda(s) \, ds, \quad t \geq 0$$

is called the mean-value function. The following result is established. $N(t+s) - N(t)$ is a Poisson random variable with mean $m(t+s) - m(t)$

The quantity $\lambda(t)$ is considered the intensity of time t or the likelihood of that event occurring around that time.

2.10 Conditional Expectation and Conditional Variance

If X and Y are jointly discrete random variables, we define $E[X|Y = y]$, the conditional expectation of X given that $Y = y$, by

$$E[X|Y = y] = \frac{\sum_x P\{X = x, Y = y\}}{P\{Y = y\}}$$

In other words, the conditional expectation of X , given that $Y = y$, is defined like $E[X]$ as a weighted average of all the possible values of X , but now with the weight given to the value x being equal to the conditional probability that X equals x given that Y equals y .

We can also find the conditional expectation with the following equation

$$E[X|Y = y] = \frac{\int x f(x, y) dx}{\int f(x, y) dx}$$

It is important to also note

$$[E[E[X|Y]] = E[X], \quad \text{and} \quad \text{Var}(X|Y) = E[X^2|Y] - (E[X|Y])^2$$

It can lastly be said that

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$