ELEC/STAT 321

Stochastic Signals and Systems

Charles Clayton

February 4, 2017

'

Table of Contents

List of Figures

List of Tables

List of Equations

1   Probability

A **random experiment** is when an outcome cannot be determined *exactly* beforehand. The **sample space**, Ω, is all the possible outcomes of the random experiment. So you might not know whether or not the alcohol content of a beer your brewing will be 5% or 6%, but you know $\Omega = [0, 100\%]$.

Subsets of Ω are called **events**, denoted as A, B, C, etc. An event **occurs** if the outcome, ω, is in the event. That is, $\omega \in A$. For instance, in our beer-brewing example, the event could be that $A = [3, 9\%]$. If the outcome is 5%, then A occurs.

A **sigma-field**, $\mathcal{F}$, is a collection of events that we can assign probabilities to. At minimum, the sample space and the empty-set must be elements of the sigma-field.

$$\emptyset, \Omega \in \mathcal{F}$$

If the event A is in the sigma field, then $\overline{A}$ must also be in the sigma field.

$$A \in \mathcal{F} \implies \overline{A} \in \mathcal{F}$$

The sigma-field is the domain of the **probability function**, P, because P transforms these collections of events into probabilities.

$$P : \mathcal{F} \to [0, 1]$$

Note that the probability function is not like other functions, the input is not a number it is a set of events. The probability of the entire sample space is 1.

$$P(\Omega) = 1$$

For disjoint events A_n, the probability of the union of the events will be the sum of the probabilities of each event.

$$A_n \text{ disjoint} \implies P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$$

1.1   Some results to use

$$P(\overline{A}) = 1 - P(A) \tag{1}$$

Probability of complement

$$A \subset B \implies P(A) \leq P(B) \tag{2}$$

Probability of subset

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{3}$$

Probability of union

Intuitively, (3) makes sense because if you have two Venn diagrams that overlap somewhat, to determine the area of the entire area, you add the two circles A and B together, then subtract the overlapping area $(A \cap B)$ because when you add A and B you count that area twice.

$$P(\bigcup_{n=1}^{n} A_n) \leq \sum_{n=1}^{n} P(A_n) \tag{4}$$

Boole's inequality

Recall that where the events A_n are disjoint, then the inequality becomes an equality.

$$P(A) = \frac{\text{\# of outcomes in } A}{\text{\# of outcomes in } \Omega} \tag{5}$$

Probability of equally likely outcomes

The number of ways you can choose k items from a set of n items.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{6}$$

Combinatorial formula

Example

Consider the 6/49 lottery. There is a box with 50 balls numbers 0-49, and 6 balls are chosen at random. What are the odds that of the 6 numbers you choose, x of those numbers are the picked balls.

—

2

The numerator is the number of possible combinations of draws that give you exactly x matches. If you are matching x numbers, then you are not matching $6 - x$ numbers. So of 6 you will have to choose x, and of the remaining 44 you will have chosen $6 - x$.

The denominator is then the number of all possible ways you can draw numbers.

$$p(x) = \frac{\binom{6}{x}\binom{44}{6-x}}{\binom{50}{6}}$$

So the probability of getting all 6 correct is $1/15{,}890{,}700 = 0.0000000063$

2 Conditional Probability

The idea here is that although the outcome can be any element in the sample space, Ω, the range of possible outcomes can be narrowed down with the help of partial information, called a **conditioning event**. For instance, if we know that the beer we brewed is a weak beer, this is a conditioning event and we can narrow down the outcome from $[1 - 4\%]$.

Where $P(A)$ is the event of interest and B is the partial information.

$$P(A|B) = P(A) \text{ given } B = \frac{P(A \cap B)}{P(B)} \tag{7}$$

$$P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B) \tag{8}$$

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B) \tag{9}$$

$$P(A|B) = 1 - P(\overline{A}|B) \tag{10}$$

Conditional probabilities

Our previous results still hold with conditional probabilities:

$$P(\Omega|B) = 1$$

$$P(A|B) \geq 0$$

If A_n are disjoint, then

$$P\left(\bigcup A_n|B\right) = \sum P(A_n|B)$$

3

Example

Without any information, the probability that you get an A in your class is the probability that your grade, $g = [80, 100]$ out of possibilities $[0, 100]$.

$$P(\text{Getting an A}) = \frac{P([80, 100])}{P([0, 100])} = \frac{2}{10} = 0.2$$

However with the partial information that you will get at least a B, so $g \geq 70$, we can use that as our conditioning event.

$$= \frac{P(A \cap B)}{P(B)} = \frac{P([80, 100] \cap [70, 100])}{P([70, 100])} = \frac{P([80, 100])}{P([70, 100])} = \frac{2}{3} = 0.667$$

2.1 Screening Tests

Conditional probability can be used for screening tests. If the test is positive, the condition is present. Consider testing manufactured items for defects. The item can be defective, D, or non-defective $\overline{D}$.

If an item is defective, the probability that the test will sense that (test positive) is the **sensitivity** of the test. This is how often your test will sense the defect.

$$\text{Sensitivity} = P(B|D) \tag{11}$$

If an item is non-defective, the probability that the test will be negative is called the **specificity**.

$$\text{Specificity} = (P(\overline{B}|\overline{D}) \tag{12}$$

Example

Say the sensitivity of your test is 0.95, the specificity is 0.99, and the probability of products actually being defective is $P(D) = 0.02$. What is the probability that the test will be incorrect?

—

Where D indicates the item is defective, and B indicates the test is positive:

$$P(B|D) = 0.95, \ P(\overline{B}|D) = 1 - 0.95 = 0.05$$

$$P(\overline{B}|\overline{D}) = 0.99, \ P(B|\overline{D}) = 1 - 0.99 = 0.01$$

The probability of the test testing positive:

$$P(B) = P(B \cap D) + P(B \cap \overline{D}) = P(D)P(B|D) + P(\overline{D})P(B|\overline{D})$$

$$= (0.02)0.95 + (1 - 0.02)0.01 = 0.0288$$

Probability of item being defective if the test is positive (accurate positive):

$$P(D|B) = \frac{P(B \cap D)}{P(B)} = \frac{(0.02)0.95}{0.0288} = 0.6597$$

Probability of item being defective if the test is negative (accurate negative):

$$P(D|\overline{B}) = \frac{P(\overline{B} \cap D)}{P(\overline{B})} = \frac{P(D)P(\overline{B}|D)}{P(\overline{B})} = \frac{(0.02)0.05}{1 - 0.0288} = 0.00103$$

The probability of the test being incorrect (false positive + false negative):

$$P(B \cap \overline{D}) + P(\overline{B} \cap D) = P(\overline{D})P(B|\overline{D}) + P(D)P(\overline{B}|D) = (0.98)0.01 + (0.02)0.05 = \mathbf{0.0108}$$

—

The screening test formula can summarized with the Bayes' Formula.

$$P(D|B) = \frac{P(D \cap B)}{P(B)} = \frac{P(B|D)P(D)}{P(B|D)P(D) + P(B|\overline{D})P(\overline{D})} \qquad (13)$$

Simple Form Bayes' Formula

Generalized, the formula looks like so:

$$P(D_i|B) = \frac{P(D_i \cap B)}{P(B)} = \frac{P(B|D_i)P(D_i)}{\sum_{j=1}^{k} P(B|D_j)P(D_j)} \qquad (14)$$

General Form Bayes' Formula

Where D_1, D_2 ... D_k is a **partition** of sample space Ω, meaning each D_i is disjoint ($D_i \cap D_j = \emptyset$ for $i \neq j$) and each outcome in the sample space is accounted for ($D_1 \cup D_2 \cup \cdots \cup D_k = \Omega$).

Let's use this iterative formula with another example.

Example

The items to be tested have two components, c_1 and c_2, say an IC wafer and a package. Suppose

$D_1 = \{\text{Only component } c_1 \text{ is defective}\}$, $P(D_1) = 0.01$

$D_2 = \{\text{Only component } c_2 \text{ is defective}\}$, $P(D_2) = 0.008$

$D_3 = \{\text{Both components are defective}\}$, $P(D_3) = 0.002$

$D_4 = \{\text{Both components are non-defective}\}$, $P(D_4) = 0.98$

$B = \{\text{Screening test is positive}\}$

And $P(B|D_1) = 0.95$, $P(B|D_2) = 0.96$, $P(B|D_3) = 0.99$, $P(B|D_4) = 0.01$

What is the probability of testing positive? What is the probability component c_1 is defective when the test is positive? c_2? What is the probability that the item is defective when the test results negative? What is the probability both components are defective when the test results positive? What's the probability of a testing error?

—

Probability of testing positive:

$$P(B) = \sum_{i=1}^{4} P(B|D_i)P(D_i) = 0.95(0.01) + 0.96(0.008) + 0.99(0.002) + 0.01(0.98) = \mathbf{0.02896}$$

Note that the probability of defects is:

$$P(D) = P(D_1) + P(D_2) + P(D_3) = 0.01 + 0.008 + 0.002 = 0.02$$

So our test will have false positives.

Probability of defect in only c_1 is:

$$P(D_1|B) = \frac{P(D_1 \cap B)}{P(B)} = \frac{P(B|D_1)P(D_1)}{P(B)} = \frac{0.95(0.01)}{0.02896} = \mathbf{0.32804}$$

6

Probability of defect in only c_2:

$$P(D_2|B) = \frac{P(D_2 \cap B)}{P(B)} = \frac{P(B|D_2)P(D_2)}{P(B)} = \frac{0.96(0.008)}{0.02896} = \mathbf{0.26519}$$

Probability of defect in both c_1 and c_2:

$$P(D_3|B) = \frac{P(D_3 \cap B)}{P(B)} = \frac{P(B|D_3)P(D_3)}{P(B)} = \frac{0.99(0.002)}{0.02896} = \mathbf{0.06837}$$

Probability neither c_1 or c_2 are defective when the test is positive:

$$P(D_4|B) = \frac{P(D_4 \cap B)}{P(B)} = \frac{P(B|D_4)P(D_4)}{P(B)} = \frac{0.01(0.98)}{0.02896} = \mathbf{0.33840}$$

Redo substituting B with $\overline{B}$ to determine the probabilities when the test is negative.

The probability of error is then:

$$P(D_4|B) + P(D_1|\overline{B}) + P(D_2|\overline{B}) + P(D_3|\overline{B})$$

2.2 Independence

When A and B are independent events, the probability of A intersect B is the product of their probabilities:

$$A \text{ and } B \text{ are independent} \iff P(A \cap B) = P(A)P(B)$$

Condition for independence

If A and B are independent, then:

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

Conditional probabilities with independence

Note that if events A and B are disjoint (mutually exclusive) or A is a true subset of B then they are dependent:

$$A \cap B = \emptyset \implies A \text{ and } B \text{ are not independent}$$

$$A \subset B \implies A \text{ and } B \text{ are not independent}$$

So in order to be independent, A and B must have some overlap, but one must

not be completely within the other.

3 Random Variables

Example

Suppose that X is Unif(0,1). Derive the distribution function and density function for $Y = -\ln(X)$.

—

Since X is Unif(0,1), then:

$$f_X(x) = 1, \ 0 \le x \le 1$$

$$F_X(x) = x, \ 0 \le x \le 1$$

Notice that the range of Y is $(0, \infty)$, so for $y < 0$:

$$F_Y(y) = 0$$

For $y > 0$:

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(-\ln(X) \le y) \\
&= P(X \ge e^{-y}) \\
&= 1 - F_X(e^{-y}) \\
&= 1 - e^{-y}
\end{aligned}
\tag{15}
$$

Similarly, for $y > 0$:

$$f_Y(y) = F_Y'(y) = \frac{\partial}{\partial y}(1 - e^{-y}) = e^{-y}$$

—

4 Random Vectors

Appendix

A Tutorial 1 (Lena)

B Tutorial 2 (Yang)

C Tutorial 3 (Lena)

C.1 Random Variables Revisited

Usually notated with an upper case letter, say X. The thing that differentiates it from an unknown x is that it can change on the fly.

It is based on the probability density function, PDF, $f_X(x)$, where the subscript is the random variable and the input is the x-axis.

Densities can take values above 1, it is not a probability, it is a measure of relative frequency. We only use this when we're plotting or integrating. Usually it's a pain otherwise.

The cumulative distribution function, CDF, $F_X(x)$ is the probability that our input variable $x \leq X$. Cumulative because it sums PDF.

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{\infty} f_x(t)dt$$

The **quantile**, q, is such a value of x that is the inverse of the CDF. Where p is the area under the $f_X(x)$ curve below $x = q$.

$$P(X \leq q) = p$$

1st quantile median.

$$P(X \leq q) = 0.25$$

$$P(X \leq q) = 0.5 = F_x(q)$$

10

Properties of CDFs and PDFs:

1. PDF

   (a) $f_X(x) \geq 0$

   (b) if $a \leq x \leq B$, then $f_X(x) = 0$ for $x \in (-\infty, a] \cup [B, \infty)$

   (c) $\int_{-\infty}^{\infty} f_X(x) = 1$

2. CDF

   (a) $0 \leq F_X(x) \leq 1$

   (b) if $a \leq x \leq B$,then
       $$F_X(B) = 1 = P(X \leq B) \text{ and } F_X(a) = 0 = P(X \leq a)$$

   (c) $F_X(x)$ is non-decreasing in x

C.2 Uniform Distribution

When the PDF is plotted, looks like a rectangle.

$$f_X(x) = \frac{1}{b-a}, \; a \leq x \leq b$$

$$F_X(x) = \frac{x-a}{b-a}, \; a \leq x \leq b$$

Since the probability is the same everywhere, if you integrate you'll see that the expected value is just in the middle.

$$E(X) = \frac{a+b}{2}$$

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

C.3 Exponential Distribution

When the PDF is plotted, it looks like a typical delay.

Usually used to model time between events. There is only one parameter λ, which is the rate per item of time or per space. X is then measured in units of whatever λ is.

$X \sim Exp(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0$$

$$F_X(x) = 1 - e^{-\lambda x}, x \geq 0$$

$$E(x) = \frac{1}{\lambda}$$

Example

$\lambda = 3/day.$ $X =$ time until next call. So $X \sim Exp(3)$.

What is the probability that the next call is within 5-6 hours?

$$P(5 \ hours \leq X \leq 6 \ hours)$$

$$P(5/24 \ days \leq X \leq 6/24 \ days)$$

$$= \int_{5/24}^{6/24} 3e^{-3x} dx = F(6/24) - F(5/24) = (1 - e^{-3(6/24)}) - (1 - e^{-3(5/24)})$$

What is the median time until the call? This means when 50% of the time is below and 50% of the time is above.

$$F_X(m) = 0.5 = 1 - e^{-3m} \rightarrow \text{solve for } m$$

Expected time for a call is $\frac{1}{\lambda}$. Variance is $\left(1/\frac{3}{24}\right)^2$

C.4 Normal Distribution

$f_X(x)$ looks like a bell-curve, symmetric around μ. Parameters are μ, the mean, and σ^2, the variance.

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(t-\mu)^2}{2\sigma^2}}$$

$$F_X(x) = \cdots = \Phi(x)$$

We can't evaluate the CDF in closed-form integration.

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

Specific case we'll often see is **standard normal**, Z. At which $Z \sim N(0,1)$, $\mu = 0$ and $\sigma^2 = 1$.

$$X \sim \mu(\mu, \sigma^2)$$

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

$$\Phi(-c) = 1 - \Phi(c)$$

$$P(|Z| \leq c) = P(-c \leq Z \leq c) = \Phi(c) - \Phi(-c) = 2\Phi(c) - 1$$

Since we can't integrate in closed-form, we'll use R functions.

CDF can be found using $\Phi(c) = $ `pnorm(c)`.

$$\Phi(2) = P(Z \leq 2) = 0.9772$$

This is the area under the curve until the point $x = 2$.

Note: Symmetry means the median is the same as the mean.

For quantities, $\Phi(c) = 0.75$, where $c = $ `qnorm(0.75)` $= 0.674$. This means that the area under the curve at the point $x = 0.674$ is 75% of the area under

the curve.

Problem B.5

$Z \sim N(0,1)$, $\Phi(1) = 0.3413$.

1. (a) $P(-1 < Z < 1)$

$$P(-1 < Z < 1) = P(Z < 1) - P(Z < -1) = \Phi(1) - \Phi(-1)$$

$$= \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 = 0.6826$$

   (b) $P(Z > 1)$

$$P(Z > 1) = 1 - P(Z \leq 1) = 1 - \Phi(1) = 0.1587$$

   (c) $P(Z < -1)$

$$P(Z < -1) = \text{same as above by symmetry}$$

2. (a) Note there are typos here, the $P(Z < c)$ should say c_1 and c_2 respectively. c_1 is larger.

$$\Phi(c_2) = 0.8, \; c_2 = \texttt{qnorm(0.8)} = 0.84$$

$$P(|Z| < c_1) = 2\Phi(c_1) - 1 = 0.8$$

$$\Phi(c_1) = 1.8/2 = 0.9$$

$$c_1 = \texttt{qnorm(0.9)} = 1.28$$

—

Non-standard normal.
$$X \sim N(\mu, \sigma^2)$$

$X = \mu + \sigma Z$, where $Z \sim N(0,1)$

$$P(X \leq c) = P(\frac{x - \mu}{\sigma} \leq \frac{c - \mu}{\sigma}) = P(Z \leq \frac{c - \mu}{\sigma}) = \Phi(\frac{c - \mu}{\sigma})$$

Properties $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$

- $X_1 \pm X_2, \sim N(\mu \pm \mu_2, \sigma_1^2 + \sigma_2^2)$

- $aX_1 \pm BX_2, \sim N(a\mu \pm B\mu_2, a^2\sigma_1^2 + B^2\sigma_2^2)$

  Important distinction

$$Var(aX_1) = a^2 Var(X_1) = a^2\sigma^2$$

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) = \sigma_1^2 + \sigma_2^2$$

Example

Distribution of grades is normal mean 75 and standard deviation is 10. Define X as the test score of the student.

1. What is the probability that a random student passes?

$$X \sim N(75, 10^2)$$

$$P(\frac{X - 75}{10} \geq \frac{50 - 75}{10})$$

$$P(Z \geq -2.5)$$

$$1 - P(Z < -2.5)$$

$$1 - \Phi(-2.5) = \texttt{1 - pnorm(-2.5)} = 0.993709$$

2. What is the P that the student scored higher than average but lower than 85?

$$P(75 \leq X \leq 85) = P(\frac{75 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{85 - \mu}{\sigma})$$

$$= P(\frac{75 - 75}{10} \leq Z \leq \frac{85 - 75}{10}$$

15

$$= P(0 \leq Z \leq 1) = \Phi(1) - \Phi(0) = \texttt{pnorm(1)} - \texttt{0.5} = 0.3413$$

Problem 3, Lecture 4

A machine fills "10-pound bags" of concrete mix. The actual weight is normally distributed with standard deviation of $\sigma = 0.1$ lbs. The mean is set by operator. Let X by the weight of the bag, where

$$X \sim N(\mu, 0.1^2)$$

1. What μ do we need if at most 10% of bags should be underfilled (< 10lbs)?

$$P(X < 10) \leq 0.10$$

   Standardize, subtract the mean.

$$P(\frac{X - \mu}{\sigma} < \frac{10 - \mu}{\sigma}) = P(Z < \frac{10 - \mu}{0.1}) \leq 0.1$$

   Solve $\Phi(\frac{10-\mu}{0.1}) = 0.1$

$$\texttt{qnorm(0.1)} = -1.2316 = \frac{10 - \mu}{0.1} \rightarrow \mu = 10.128$$

   If you increase μ, you decrease the probability that the $X < 10\%$ because you reduce the area under the curve below 0.1.

   So for $P(X < 10) \leq 0.1$, we need $\mu \geq 10.128$.

2. Now define $\sigma = 0.1\mu$. Hence we have $X \sim N(\mu, 0.1^2 \sigma^2)$. Same question as above.

   Standardize again:

$$P(X < 10) = P\left(\frac{X - \mu}{\sigma} < \frac{10 - \mu}{\sigma}\right) = P\left(Z < \frac{10 - \mu}{0.1\mu}\right) = \Phi\left(\frac{10 - \mu}{0.1\mu}\right)$$

$$\frac{10 - \mu}{0.1\mu} = -1.216(???) \rightarrow \mu \geq 11.47$$

C.5   Failure Rates

Easy problems, just know the definitions of CDF and PDF.

Example

Lifetime of relationship has failure rate in years:

$$\lambda(t) = \frac{1}{4}e^{-t}$$

1. Are relationships stronger or weaker over time?

   Since the failure rate decreases over time, then the chance that the relationship breaks is shorter the longer the relationship is, so relationships get stronger over time.

2. Mike has just started to date Lindsey. What is the probability that their relationship survives 6-months. X = duration of relationship.

   $$F_X(x) = 1 - \exp \int_0^x \lambda(t)dt$$

   $$P(X > 0.5) = 1 - P(X \le 0.5) = 1 - F_X(0.5) = 1 - (1 - \exp(-\int_0^{0.5} \frac{1}{4}e^{-t}dt))$$

   $$= \exp(-\int_0^{0.5} \frac{1}{4}e^{-t}dt) = \exp(\frac{1}{4}e^{-0.5} - \frac{1}{4}) = 0.90631$$

   We solved this on the fly. Don't take it as gospel.

3. Find the median lifetime $F_X(m) = 0.5$.

   $$1 - \exp(-\int_0^m \lambda(t)dt) = 0.5$$