Part 1: Logic and Probability

In some sense, probability subsumes logic: While a probability can be seen as a measure of “degree of truth”—a real number between 0 and 1—logic deals merely with the two extreme values, 0 (false) and 1 (true).

This statement is not quite rigorous: An event of probability zero is not logically impossible, it is just extremely unlikely to occur. Similarly, an event of probability one is not certain to occur, but rather extremely likely to occur. As an example, imagine making a measurement (a voltage, a time, ...) that returns a number $x$ in the set $\mathbb{R}$ of reals. Before making the measurement, and in the absence of prior information, most reasonable models would assign probability zero to the event $X = \{x\}$. After all, what are the chances that we guess the value of $x$ exactly right among all real numbers? This implies that the complementary event $Y = \mathbb{R} \setminus X$ has probability 1. An yet $X$ occurs and $Y$ does not.

With this caveat, a probability of 1 is as close as probability theory gets to certainty, and a probability of 0 to impossibility.

Problem 1.1 (tag: contrapositive)

Given logical propositions $a$ and $b$, the contrapositive of the conditional $a \rightarrow b$ is the conditional $\sim b \rightarrow \sim a$. A natural translation of $a \rightarrow b$ into the language of probabilities is to say that the probability that $b$ is true given that $a$ is true is equal to 1.

Formally, consider the sample spaces

$$S_a = S_b = \{T, F\}$$

where $T$ stands for “true” and $F$ for “false.” A compound experiment on $a$ and $b$ then has compound sample space

$$\mathbb{S} = S_a \times S_b = \{(T,T), (T,F), (F,T), (F,F)\} .$$

The events

$$A_T = \{(T,T), (T,F)\} \text{ and } B_T = \{(T,T), (F,T)\}$$

then mean, respectively, that $a$ is true (regardless of the value of $b$) and that $b$ is true (regardless of the value of $a$), and the conditional $a \rightarrow b$ can be rephrased as follows:

$$\text{prob}(B_T | A_T) = 1 .$$

Use the rules of the calculus of probability to prove that

$$\text{prob}(B_T | A_T) = 1 \text{ implies } \text{prob}(A_F | B_F) = 1$$

where

$$A_F = \{(F,T), (F,F)\} \text{ and } B_F = \{(T,F), (F,F)\}$$

mean that $a$ is false and $b$ is false, respectively. This implication is a probabilistic rephrasing of the equivalence of a conditional with its contrapositive.

Show your derivation in detail. That is, prove any fact that you need and that was not proven in class or in the required readings. If you use a previously proven result, mention that result explicitly.

[Hint: First prove in detail that $\text{prob}(A_T | B_F) + \text{prob}(A_F | B_F) = 1$.]
Part 2: Forward and Inverse Probabilities

A probabilistic experiment is a way to generate random outcomes. As an example running through this part of the assignment, consider drawing a ball out of one of $U$ urns: For $u = 1, \ldots, U$, and $c = 1, \ldots, C$, urn $u$ has $n_{uc}$ balls of color $c$, which weigh $W(c)$ grams each. Here, $W(c)$ is a random variable defined on the sample space of all colors,

$$
S_c = \{1, \ldots, C\}.
$$

In the experiment, we first pick a value for $u$ at random, with probability $P(u)$. Then we pick a ball out of urn $u$. The quantities $U, C, n_{uc}, P(u), W(c)$ are parameters of the experiment.

A forward probabilistic calculation assumes that all the parameters are known, and computes a probability or an expectation based on the outcome of the experiment, such as the probability of drawing a green ball or the expected weight of the ball being drawn. An inverse probabilistic calculation assumes that only some of the parameters are known. Given the outcome of some experiment, possibly repeated, the inverse calculation returns the value of the unknown parameters. This part explores both forward and inverse probabilistic calculations.

For each question in this part, give both a general formula (or formulas) in terms of the parameters given above, and a numerical, decimal value (or set of values). Use the following numerical values, except when a problem specifies different values:

$$
U = 2 \quad , \quad C = 3
$$

$$
P(1) = 0.3 \quad , \quad P(2) = 0.7
$$

$$
w_1 = 200 \quad , \quad w_2 = 100 \quad , \quad w_3 = 300
$$

$$
n_{11} = 3 \quad , \quad n_{12} = 5 \quad , \quad n_{13} = 4
$$

$$
n_{21} = 7 \quad , \quad n_{22} = 0 \quad , \quad n_{23} = 6
$$

Round decimal values to three decimal digits after the period.

**Notation:** In keeping with standard convention, we use the same symbol $P$ for all probability functions, and use the argument to disambiguate what the function stands for. For instance, $P(u)$ is the probability of drawing from urn $u$, while $P(c)$ is the probability that a ball has color $c$. As discussed in class, this notation causes difficulties when variables are replaced with numbers. For the parameter values $P(u)$ listed above, for instance, context implies that the integers refer to the urn number.

**Hint:** It may save you both time and mistakes to do the calculations in this part by a (carefully checked) piece of code, rather than by hand or calculator. However, do not show your code.

**Problem 2.1 (tag: forward)**

Give the following probabilities. The first one is done for you so you see how to format your answers.

- The conditional probability function $P(c \mid u)$ that a ball of color $c$ is drawn from urn $u$ (that is, given that the chosen urn is $u$). Arrange the numerical values into a $U \times C$ array.

- The joint probability function $P(u, c)$ that a ball is drawn from urn $u$ and has color $c$, in terms of $P(c \mid u)$ and $P(u)$. Arrange the numerical values into a $U \times C$ array.

- The probability function $P(c)$ that a ball drawn as described in the introduction has color $c$, in terms of $P(u, c)$. Arrange the numerical values into a $1 \times C$ array.

- The probability function $P(u \mid c)$ that the ball was drawn from urn $u$, given that its color is $c$. Write your formula in terms of $P(c \mid u)$ and $P(u)$ (do not use $P(c)$ directly in your expression). Arrange the numerical values into a $U \times C$ array.

$$
P(c \mid u) = \frac{n_{uc}}{\sum_{k=1} C n_{uk}} = \begin{bmatrix}
0.250 & 0.417 & 0.333 \\
0.538 & 0.000 & 0.462
\end{bmatrix}
$$

**Problem 2.2 (tag: expectations)**

Give the mean $m_W$ and standard deviation $\sigma_W$ of the weight of a ball drawn as described in the introduction. Include units of measure.
Problem 2.3 (tag: inverse)

Inverse problems use the same set of equations as the forward calculations to solve for different unknowns.

Suppose that there are $U = 2$ urns, and that the probability function $P(u)$ is unknown. All other quantities have the values given in the introduction. Based on a large number of draws, it is established that 30 percent of all balls drawn have color $c = 1$.

Give the new distributions $P(u)$ and $P(c)$. Show your reasoning.

**Notation:** The notation adopted so far in this problem becomes confusing when numbers are replaced for letters. To avoid confusion, write $c_k$ or $u_k$ rather than just $k$. For instance, the value of the distribution $P(c \mid u)$ for $c = 1$ and $u = 2$ is denoted by $P(c_1 \mid u_2)$. You may also define new variables to streamline your calculations. For instance,

$$ p = P(u_1) .$$
Part 3: Camera Pixels

A cellphone camera has millions of pixels in it. Each pixel is a device that counts photons of light. To model how this happens, let us discretize time into intervals that are $\tau = 0.1$ milliseconds in length. In any given interval, either no photon or one photon arrives at the pixel. If photon arrivals are independent, they can be described as a sequence of repeated Bernoulli trials, with “success” in trial $k$ corresponding to the arrival of a photon in interval $k$. The success probability $p$ for an individual arrival in any interval of length $\tau$ increases with the brightness in the scene. The assumption that at most one photon arrives at the pixel in a time interval of length $\tau$ implies that the value of $p$ is very small, so the analysis in terms of Bernoulli trials holds only for relatively dark scenes.

With probability $\eta$, a photon that arrives at a pixel dislodges an electron from a semiconductor crystal inside the pixel. Whether one photon dislodges an electron is independent of whether a different photon does. The dislodged electron is captured by an electric field and stored in a capacitor inside the pixel. Thus, the arrival of electrons at the capacitor is a repeated Bernoulli trial with probability $p_e = \eta p$. The probability $\eta$ is called the pixel’s quantum efficiency, and is typically very high (0.8 or more).

To take a picture, the pixel first discharges the capacitor, so it has no free electrons in it. Photons start to flow in and produce electrons. After an exposure time $T = n\tau$, where $n$ is an integer in our analysis, an electronic circuit counts the electrons and reports their number. The brighter the part of the scene viewed by that pixel, the higher the electron count.

Repeated Bernoulli trials and related probability distributions are discussed in the April 13 class notes and in the textbook (pages 370-375 are part of the required readings).

Problem 3.1 (tag: dark1)

Let the exposure time $T$ for a camera be 30 milliseconds, that is, $T = n\tau$ with $n = 300$, and let $\eta = 0.9$ be the camera’s quantum efficiency. A pixel is dark if no electrons arrive at the capacitor over the exposure period $T$. If the scene brightness corresponds to a photon arrival probability $p = 10^{-2}$, what is the probability $p_d$ that a pixel is dark? Give both a formula and a numerical value in scientific notation (example of scientific notation: $2.345 \cdot 10^{-4}$).

---

1 Thus, we assume that brightness levels are low enough that it is virtually impossible for two photons to arrive in the same time interval.

2 In the form of a voltage, but this is irrelevant.
Part 4: Poisson Arrivals

[This part involves differential equations, but do not worry. You will only need some basic facts from calculus.]

The calculation performed in the previous part is based on an approximate model of electron arrivals that discretizes time. That is, the model allows at most one photon (or η electrons) to arrive every tenth of a millisecond. A question then arises as to whether this approximation leads to significant errors in any estimates that are based on it. This part derives an exact model without time discretization, and the calculation we did in the previous part is repeated in this model.

The exact model is called a Poisson arrival model, and is very important in many branches of science and mathematics. The derivation of this model is therefore of great conceptual importance, even if we use it only in a small way in this assignment. The derivation is also a good opportunity to review probability calculations and proofs by induction.

Let $P(n, δt)$ be the probability that there are $n \geq 0$ arrivals in a time interval $δt \geq 0$. We make the following assumptions:

1. The probability of one arrival in a very short time interval is proportional to the length of the interval:

$$\lim_{δt \to 0} \frac{P(1, δt)}{δt} = \lambda .$$

where the positive real number $\lambda$ is called the arrival rate and has the units of number of arrivals per second.

2. The probability $1 - P(0, δt) - P(1, δt)$ of two or more arrivals vanishes as $δt$ goes to zero:

$$\lim_{δt \to 0} P(0, δt) + P(1, δt) = 1 .$$

3. The numbers of arrivals in disjoint time intervals are independent.

Because of assumption 1, the probability that there is one arrival in a very short interval $δt$ is approximately

$$P(1, δt) \approx λδt ,$$

and because of the first two assumptions, the probability that there is no arrival in a very short interval is approximately

$$P(0, δt) \approx 1 - λδt .$$

These approximations become better and better as $δt \to 0$.

**Problem 4.1 (tag: basics)**

Answer the following preliminary questions about the Poisson arrival model:

- Why does the proportionality to $δt$ only hold for short intervals? In other words, why can we not say

  $$P(1, δt) = λδt \quad \text{for any } δt \geq 0$$

  in assumption 1?

- What is the numerical value of $P(1, 0)$?

- What is the numerical value of $P(0, 0)$?

**Problem 4.2 (tag: p0t)**

Determine the relationship between $P(0, t)$ and $P(0, t + δt)$ where $δt$ is a small enough but not zero interval.

To this end, subdivide the interval between time 0 and time $t + δt$ into two subintervals: The first extends from time 0 to time $t$, and the second covers the remaining part, from time $t$ to time $t + δt$.

Use independence (assumption 3) and the fact that $δt$ is small enough for equations (1) and (2) to hold and write $P(0, t + δt)$ in terms of $P(0, t), \lambda$, and $δt$. 

Spring 2017
Problem 4.3 (tag: diff0)

If you answered the previous problem correctly, it should now be possible for you to rearrange terms in the equation you found, then apply the definition of derivative

\[
\frac{df}{dt} = \lim_{\delta t \to 0} \frac{f(t + \delta t) - f(t)}{\delta t}
\]

to show that

\[
\frac{dP(0, t)}{dt} = -\lambda P(0, t).
\]

This is a differential equation in \( P(0, t) \).

Derive this equation as instructed above, showing all your steps.

Problem 4.4 (tag: diff1)

We now find a differential equation in \( P(n, t) \) for \( n > 0 \), using similar methods as in the previous problem.

We again subdivide the interval between time 0 and time \( t + \delta t \) into two subintervals, as we did for \( P(0, t) \). Because of property 2, the probability of having more than two arrivals in the second subinterval vanishes as \( \delta t \) goes to zero. The outcome of \( n \) arrivals in the whole interval is therefore the disjunction of two mutually exclusive cases:

- \( n \) arrivals between time 0 and time \( t \) and no arrivals between time \( t \) and time \( t + \delta \); or
- \( n - 1 \) arrivals between time 0 and time \( t \) and one arrival between time \( t \) and time \( t + \delta \).

Use this observation to write an expression for \( P(n + \delta t, t) \) in terms of \( P(n, t) \), \( P(n - 1, t) \), \( \lambda \), and \( \delta t \) for a value of \( \delta t \) that is small enough that equations (1) and (2) hold.

Then, rearrange the terms in your equation and take the limit for \( \delta t \to 0 \), similarly to what you did for \( P(0, t) \), to obtain the following differential equation for \( P(n, t) \):

\[
\frac{dP(n, t)}{dt} = \lambda [P(n - 1, t) - P(n, t)].
\]

You are given the end result, just show the steps to get there.

Summary So Far

We now found two differential equations, one for \( P(0, t) \) and one for \( P(n, t) \). Combining them into one, we have

\[
\frac{dP(n, t)}{dt} = \begin{cases} 
-\lambda P(0, t) & \text{if } n = 0 \\
\lambda [P(n - 1, t) - P(n, t)] & \text{otherwise.}
\end{cases}
\] (3)

In the following two problems, you will show by induction on \( n \) that the solution to these two equations is

\[
P(n, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad \text{for } n \geq 0.
\] (4)

This is called the Poisson probability distribution with arrival rate \( \lambda \).

Caveat: It is a bit unfortunate that \( P \) here is a function, because you may be used to thinking of \( P \) as the predicate that is to be proven by induction. To avoid confusion, you may want to use \( Q \) to denote the predicate instead. What you are proving is not that

\[
P(n, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad \text{for } n \geq 0.
\]

(What would it mean to prove this expression, anyway? This equation merely defines a function \( P(n, t) \).) Instead, you are proving the following statement

\[
\forall n \geq 0 : Q(n)
\]

where \( Q(n) \) is the predicate

\[
\text{If } P(n, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \text{ then } \quad \frac{dP(n, t)}{dt} = \begin{cases} 
-\lambda P(0, t) & \text{if } n = 0 \\
\lambda [P(n - 1, t) - P(n, t)] & \text{otherwise.}
\end{cases}
\] (5)
Problem 4.5 (tag: base)

In this problem, you will prove the base case $n = 0$, which refers to the first differential equation:

\[ \frac{dP(0,t)}{dt} = -\lambda P(0,t). \]

The integral of this equation is

\[ P(0,t) = Ce^{-\lambda t}, \]

as you may recall from calculus. What is $C$, given the value you found earlier for $P(0,0)$? Rewrite the expression for $P(0,t)$ above with the value of $C$ you found. Then replace your expression into the first differential equation above and verify that it holds.

Problem 4.6 (tag: induction)

The inductive step is straightforward for differential equations: Assume that $Q(k)$ (expression (5) with $n$ replaced by $k$) is true for an arbitrary integer $k \geq 0$. In words, assume that $P(k,t)$ (definition (4) with $n$ replaced by $k$) satisfies the differential equation (3). Can we then show that $Q(k+1)$ is true? That is, does $P(k+1,t)$ satisfy equation (3)?

Take the inductive step, and show all your reasoning.

Problem 4.7 (tag: axioms)

Show that

\[ P(n,t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad \text{for } n \geq 0. \]

is a probability distribution.

[Hints: $P(k,t)$ needs to satisfy two properties in order to be a probability distribution. What is the Taylor series of the exponential function?]

Problem 4.8 (tag: dark2)

Let us now revisit the probability of a dark camera pixel. In our original formulation, the probability of one photon arriving in 0.1 milliseconds was $p = 10^{-2}$. This corresponds to an arrival rate

\[ \lambda = \frac{\eta p}{\tau} = \frac{0.9 \cdot 10^{-2}}{10^{-4}} = 90 \quad \text{electrons per second}. \]

Assuming that the number of electrons arriving at a pixel has the Poisson distribution with this arrival rate, recompute the probability $p_d$ that a pixel is dark. Give a formula and a numerical value, assuming the same exposure period $T = 30$ milliseconds.

Was is bad to approximate Poisson arrivals with Bernoulli trials for this calculation?

Problem 4.9 (tag: dark3)

Give the numerical values of $p_d$ from both the Bernoulli model and the Poisson model with all the same parameters as above except $p = 0.2$. Is the approximation better or worse than the one obtained for $p = 10^{-2}$? Why?