Advanced Applied Probability problems
by Richard Cowan

A. Small pathologies for the unaware student

Students with a statistics, rather than mathematical, background are often interested in advanced topics in "applied probability", but are not fully aware of various complications such as distributions with no densities, with infinite expectation (and so on). These problems help their awareness.

1. Find $E(Y)$ if Y has distribution function G given as:

$$\begin{aligned} G(y) &= 0 && y < 0 \\ &= \frac{y}{3} && 0 \le y < 1 \\ &= 1 - \frac{1}{3y^2} && 1 \le y. \end{aligned}$$

   Using (a) $E(Y) = \int_{[0,\infty)} y \, dG(y)$

   (b) $E(Y) = \int_0^\infty \big[1 - G(y)\big] \, dy$

2. $X_1, X_2, \cdots, X_n$ are iid Exponential(λ) variates. The n numbers are written on separate cards and placed in a hat. One card is drawn out randomly and the number observed. Call it X. The card is *replaced* and all cards are shuffled. Another card is drawn; call the number observed Y. Write down the joint distribution function F of X and Y. Find $E(XY)$.

3. Non-negative random variables X and Y have joint distribution function F. Show that
$$\mathrm{Cov}(X, Y) = \int_0^\infty \int_0^\infty \big[F(x, y) - F_X(x)F_Y(y)\big] \, dx \, dy$$

4. (Record values) Let $X_1, X_2, \cdots$ be the winning times in the mens' 1500 metre race at successive Olympic games. Assume that these *continuous* random variables are iid (a bit unrealistic in the world of athletics). We say that X_j creates a record if it is the lowest time so far. Clearly X_1 sets a record. Define N as the number of games until this initial record is beaten. Find $p_n := P\{N = n\}$ and $E(N)$.

5. X and Y are independent, non-negative random variables with distribution functions F and G. F has a density function f but G does not have a density due to a single jump discontinuity at a. Define $Z := X + Y$.

   (a) Show that H, the distribution function of Z, is given by

$$H(z) = \int_{[0,z]} F(z-u) \, dG(u), \qquad \text{or by}$$
$$H(z) = \int_0^z f(u) \, G(z-u) \, du$$

(b) Show that H has a density (i.e. has no jump discontinuities).

B. Advanced practice problems

1. n rooms are arranged in a line with doors connecting adjacent rooms. $(n-1)$ \$1 coins are placed randomly in the rooms. Each coin is equally likely to be placed in any of the n rooms independently of other coins. A man who is in room 1 wishes to travel to room n, but it costs \$1 to open each of the doors. The man can collect any coins that he encounters during his trip, so that he has money to pay these costs. There are $n-1$ doors and $n-1$ coins, so it is <u>possible</u> for the man to complete his journey. Prove that the probability of the man reaching room n is $1/n$. [Note: Don't be deceived by the easy answer! The problem, which underpins some theory on queue busy-periods, is challenging.]

2. We are given n iid uniformly distributed variables in the interval $[0,1]$. These partition the interval into $n+1$ parts. Let X be the number of parts which are larger than b $(0 < b < 1)$. Find $\mathrm{E}(X)$.

3. A space may have many σ-rings. How many does the finite space $S = \{a, b, c, d\}$ have ? How many of them are σ-algebras? Answer the same questions for a finite space S comprising n elements. [Note: A bit out of place – an interesting counting problem drawn from my courses on Mathematical Probability.]

C. Invariant distributions in Markov processes on a continuum

These problems augment rather informal lectures on integral equation methods for establishing the invariant distribution of such Markov processes and limiting distributions (if they exist).

1. Two glasses A and B, each with capacity of 1 litre, are initially half full of water. At times $1, 2, \ldots,$ a random transfer of water from one to the other takes place. Precisely, one glass is randomly selected (probability α for A and $\beta = 1 - \alpha$ for B) and a random variable U, uniformly distributed on $[0,1]$, is sampled. Then a proportion U of the water in the chosen glass is transfered to the other glass. Find the transition probability density function $p(\cdot|\cdot)$ and set up the integral equation on the state space $[0,1]$
$$f(x) = \int_0^1 p(x|y) f(y) dy$$
which an invariant pdf f should satisfy. Assuming that this process has an equilibrium, show that the limiting distribution for the quantity of water in glass A is a Beta distribution (with considerable probability mass near the end points of [0,1]).

2. Whereas we often take the informal approach, used in problem 1, of assuming that a limiting distribution exists and is equal to the invariant distribution, we can at times use the integral recurrence on the state space A
$$f_{n+1}(x) = \int_A p(x|y) f_n(y) dy$$

to find successively the densities of $X_1,\ X_2,\ \cdots$. For example, let $U_1,\ U_2,\ \cdots$ be independent uniform variates on $[0,a]$ and define $X_n = \prod_{i=1}^{n} U_i$. Find the pdf of X_n.

3. We have to be careful not to overlook problems of existence. Consider the following problem. $X_1,\ X_2,\ \cdots$ are i.i.d. random variables and $F(x) = P\{X_i \le x\}$. Define a new sequence $\{Y_n\}$ by the relation

$$Y_n := \left(\sum_{i=1}^{n} X_i\right) \bmod 1.$$

Set up the integral equation for an invariant pdf and show that the uniform distribution is a solution. Assuming glibly that the limiting distribution exists and that it is the solution of this integral equation, we conclude that the uniform distribution is the limit. This result is true for many F, but can you find an example of F for which it is not? [Hint: What happens for lattice distributions?]

4. A problem where distribution functions should be used. $\{X_n\}_{n=1}^{\infty}$ is a sequence of independent random variables each uniformly distributed on $[0,1]$. Y_1 is defined equal to X_1 whilst Y_n is defined for $n = 2,3,\cdots$, as

$$Y_n := \max\bigl(X_n,\ \lambda Y_{n-1}\bigr)$$

for some constant λ such that $0 < \lambda < 1$. Assuming that $G_n(y) := P\{Y_n \le y\}$ has a limit $G(y)$ as $n \to \infty$, find $G(y)$. Discuss the cases $\lambda = 1,\ \lambda > 1$ and $\lambda \le 0$.

5. $\{X_n\}_{n=1}^{\infty}$ and $\{Y_n\}_{n=2}^{\infty}$ are independent random variables each uniformly distributed on $[0,1]$. Z_1 is defined equal to X_1 whilst Z_n is defined for $n = 2,3,\cdots$, as

$$Z_n = \max\bigl(X_n,\ Y_n Z_{n-1}\bigr).$$

If $G_n(z) := P\{Z_n \le z\}$, find $G_1(z)$, and show that

$$\begin{aligned} G_n(z) &= z^2 + z^2 \int_z^1 \frac{G_{n-1}(x)}{x^2}\,dx && z \in [0,1] \\ &= 0 && z < 0 \\ &= 1 && z > 1 \end{aligned}$$

Assuming that $G_n(z)$ converges for all z and that, for $z \in [0,1)$, the limit G is differentiable and is the unique solution of the equations

$$\begin{aligned} G(z) &= z^2 + z^2 \int_z^1 \frac{G(x)}{x^2}\,dx \\ G(1) &= 1. \end{aligned}$$

Find $G(z)$. [*Hint:* Let $H(z) = G(z)/z^2$ and solve for H firstly.]

6. The *shape* of a rectangle is defined by one number, namely the ratio of the shorter to longer side lengths. Thus *shape* $\in (0, 1]$. Consider the following Markov process on rectangle shapes.

   Given a rectangle R_n with one corner labelled A, we construct the next rectangle R_{n+1} in the sequence by (a) splitting R_n with a randomly positioned cut orthogonal to its long sides, thus creating two rectangular pieces, and then (b) selecting the piece which contains A. To position the cut, we place a point P on the long side such that the random variable U, defined as the length of AP, has p.d.f. g defined by

   $$g(u) = 2u \qquad 0 \leq u \leq 1$$

   and make the orthogonal cut at P.

   It can be shown that if an equilibrium p.d.f. f for shape X exists as $n \to \infty$, then it satisfies the integral equation

   $$f(x) = \int_{(0,1]} p(x|y) f(y) dy$$

   where $p(\cdot|y)$ is the conditional p.d.f. of the shape X given that the previous shape is y.

   (a) Find $p(x|y)$ and write out the integral equation in full.

   (b) Show that $f(y) = by^{b-1}$, $0 < y \leq 1$, is a solution of the equation for one particular value of b.

   (c) Find that unique value of b.

D. Point processes

1. [Due to M. Neuts] Consider a Poisson process on R of rate λ. A new point process is constructed by (a) placing one point between each pair of original points in a random manner described below, and (b) then deleting the original points. For (a), the new point is placed a proportion U along the existing interval between two points, where U has pdf. (for $0 < \alpha < 1$)

   $$f(u) = \frac{u^{\alpha-1}(1-u)^{-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)} \qquad\qquad 0 \leq u \leq 1.$$

   Each interval has an independent choice of U. Show that the new point process is a Poisson process. [Note: It can be shown that U must have a distribution of this form if one requires the new process to be Poisson (leaving aside the trivial cases where $U = 1$ or 0 with probability 1)]

2. (Adam's delay problem, 1936). Suppose that the times at which cars (with zero length) pass a certain location form a Poisson process of rate λ. A pedestrian arrives at the location at time t and desires to cross the road. He will cross as soon as there

is a gap of length $> \beta$ in the traffic. Find the *expected* delay to the pedestrian. What is the variance of delay?

3. [From Cox & Miller]. An approximate (one-dimensional) model of a textile yarn is that fibre left-ends are distributed along the yarn axis in a Poisson process of rate λ. Each fibre is of length τ and is parallel to the yarn axis. The yarn thickness at any point is the number of fibres crossing that point. Find the correlation coefficient between thicknesses at points h apart.

4. A person who invests \$1 at time t knows that it will be worth $\$e^{rx}$ at time $t+x$. Here, r is a type of 'interest rate' parameter. So income of \$1 at $t+x$ is equivalent to income of $\$e^{-rx}$ at time t since $\$e^{-rx}$ could be invested to yield \$1 at $t+x$. If we are currently at time t, we say that $\$e^{-rx}$ is the 'present value' of a \$1 payment at time $t+x$. Imagine a person who is to receive income of $\$a$ at each point of a stationary point process of intensity λ. What is the *total* present value of this 'future sequence of payments'?

5. Stationary point processes on R have 'translation invariance' of the count structure defined via the translation operator T^a, where $T^a x := a + x$. Consider another operator $I : R \to R$ defined by $Ix = -x$. This is a reflection about the origin. A point process is *reflection invariant* iff for all A_i and n_i

$$P\left\{\bigcap_{i=1}^{k}\{N(A_i) = n_i\}\right\} = P\left\{\bigcap_{i=1}^{k}\{N(IA_i) = n_i\}\right\}.$$

Find an example of a stationary process which is not *reflection invariant*. Is the stationary Poisson process reflection invariant? [*Note:* A process that is invariant under I and T^a (all a) is called *homogenous.*]

6. Consider the circle with centre $\left(0, \frac{1}{2\pi}\right)$ and radius $\frac{1}{2\pi}$. This circle, which has circumference 1, is tangential to the x-axis. Two points, A and B, are independently and uniformly distributed on the circumference of this circle. The circle is then rolled along the x-axis. Whenever the points A or B make contact with the x-axis, a new point is placed on the x-axis. Thus a stationary point process of "A and B copies" is created along the x-axis. Find the mean length of the interval that contains an arbitrary point C placed at distance t along the axis.

7. A motorist is involved in accidents occasionally. Suppose that the times at which his accidents occur can be modelled as a stationary Poisson process of intensity λ. He insures his car at time 0, but is told by the insurance company that they will stop the insurance policy as soon as he has an accident <u>within</u> a time β of an earlier accident.

   (a) Let T be the time until his insurance policy is stopped. Find $\mathrm{E}(T)$.

   (b) If N is the number of accidents before the policy is stopped (including the one which <u>causes</u> the stop), find $\mathrm{E}(N)$.

8. Consider n independent replicates of a stationary point process of intensity λ, and also their superposition. Clearly the superposition is stationary with intensity $n\lambda$. Now consider the 'rescaled' superposition formed by applying the dilation

(i.e. stretching) function $h: x \to nx$ to the superposition. Prove that, in the limit as $n \to \infty$, the distribution of the first recurrence time tends to the exponential distribution.

9. People enter a room of unlimited capacity at the epochs of a stationary Poisson process. Independently with probability θ they stay 1 hour and with probability $1-\theta$, 2 hours. Show that the exit times are also a Poisson process. More generally, prove that, if the periods of stay in the room are iid positive random variables with distribution function F, then the exit times form a stationary Poisson process.

10. Associated with each point of a stationary Poisson point process in R is a rifleman who at a given signal shoots the closest of his neighbours. Find F_1, the distribution function for the first recurrence time, for the point process of surviving riflemen. Hence find the intensity of the new point process.

11. (Olbers paradox in cosmology). Suppose that stars (treated as points) form a stationary Poisson process in R^3 with intensity θ. Assume further that stars are omitting iid amounts of energy, Y, where $E(Y) = \mu$. The energy that we observe from a star of energy y at distance ℓ from us is $\alpha y/\ell^2$, where α is a constant. Determine the expected total energy that we observe from the universe.

12. Consider the following stochastic problem concerning some impatient people in a building served by one lift.

    Assume that the point process of people arriving at the ground floor and wishing to catch the lift is a Poisson process of rate λ. On one occasion when the lift doors open at the ground floor, Jim, the only waiting passenger, enters the lift (which is empty) and immediately presses the "close door" button.

    The doors take time β to close, but if any new arrival comes before β has elapsed, this person presses a button and the doors re-open. A door which has been closing for a time $x < \beta$ will take time x to re-open. When the door re-opens, the new arrivals immediately enter and the "close door" button is pressed. The process is repeated until the door successfully closes. The lift then moves.

    Let T be the time until successful closure of the door, measured from the moment when Jim enters the lift. Find $E(T)$. [For simplicity, you may assume that the lift has no bound on capacity. Also, Jim and others entering the lift can only do so when the doors are fully open.]

13. People arrive at a ferry terminal according to a Poisson process on $[0, \infty)$ of rate λ. The ferry departs at regular intervals of a time units, at times $a,\ 2a,\ 3a,\ \cdots$. Assuming that the ferry has no upper bound on capacity, calculate the average waiting time of the people. Do the same, assuming that the ferry departure times form a general stationary point process of intensity λ.

E. Random geometry

1. A 3-4-5 pythagorean triangle with vertices $A = (0,3), B = (4,0)$ and $C = (0,0)$ is cut with a random IUR transect $\mathcal{L}$, thereby creating a new triangle and a quadrilateral.

(a) What is the probability of the event R that the newly created triangle is a right-angled triangle?

(b) Given that R occurs, the random transect cuts the sides CA and CB; we label these cutting points P and Q respectively. Define the random variables $X := |CP|$ and $Y := |CQ|$. Find the joint distribution function, $F(\cdot, \cdot|R)$ say, of X and Y conditional upon event R.

(c) Find the joint probability density function $f(x, y|R)$.

(d) Given R, find the expected length of $\mathcal{L} \cap \triangle_{ABC}$ (i.e. $E(|PQ|)$). [Note: you may quote from the table of integrals below if required.]

(e) Given R, show that the expected perimeter of the newly-created triangle is $\frac{1}{4}(25\ell n\, 5 - 16\ell n\, 2 - 10)$. [Note: Again consult the table if needed.]

Table of Useful Integrals

$$I_1 \equiv \int_0^a \sqrt{u^2 + b^2}\, du = \frac{ac}{2} + \frac{b^2}{4} \ell n \frac{c+a}{c-a} \quad (c^2 = a^2 + b^2)$$

$$I_2 \equiv \int_0^a u\, \ell n\, (b^2 + u^2) du = \frac{1}{2}[2c^2\, \ell n\, c - 2b^2\, \ell n\, b - a^2]$$

$$I_3 \equiv \int_0^a u\, \ell n\, u\, du = \frac{a^2}{4}[2\, \ell n\, a - 1]$$

2. If an equilateral triangle with side length 1 is cut by an IUR transect, show that the expected perimeter of the quadrilateral formed by the cut of a random transect is $3 + \frac{\pi\sqrt{3}}{12} - \frac{3}{4}\ell n 3$.

3. D is a convex polygon with n sides, perimeter L and area A. K is another convex polygon with n_K sides, perimeter L_K and area A_K. Consider a uniformly distributed congruent copy (K' say) of K hitting D. The domain $D \cap K'$ is itself a convex polygon. Let N be the number of sides of this polygon. Find E(N).

4. Consider a homogeneous, Poisson line process in R^2 with the property that the number of lines cutting any convex domain of perimeter 1 has expectation ρ/π.

(a) Let Z be the distance from the origin to the nearest line. Find the probability density function of Z.

(b) Let X and Y be defined as the distances, measured in a positive direction along the x-axis and y-axis respectively, until a line is encountered. Find the joint distribution function of X and Y.

(c) By using the formula

$$\text{Cov}(X, Y) = \int_0^\infty \int_0^\infty \big[F(x, y) - F_X(x)F_Y(y)\big]\, dx\, dy,$$

show that X and Y have correlation coefficient equal to

$$3 - 4\ell n 2 = 0.227.$$

[Hint: You may quote the following result without proof.

$$\int_0^{\pi/2} \frac{dx}{\left[1+\sin(x)+\cos(x)\right]^2} = 1 - \ell n 2 \quad]$$

5. Two Poisson line processes of rates ρ_1 and ρ_2 respectively are superimposed. Prove that the resulting process is a Poisson line process. For a 'typical' Miles' polygon in the first process, what is the expected number of cuts by lines from the second process?

6. Consider the set $\mathcal{V}$ of vertices of the Poisson-Voronoi tessellation. Let $\mathcal{C}_r$ be the set of all 'ortho-centres' of all triples of particles which lie in the ball B_r. Show that each $v \in \mathcal{V}$ belongs to $\mathcal{C}_r$ for a sufficiently large r. Show that the probability that any $c \in \mathcal{C}_r$ which is also $\in \mathcal{V}$ has order greater than 3 is zero. Hence conclude that all vertices in $\mathcal{V}$ are of order 3 with probability one.

7. Prove that all the polygons in the Poisson-Voronoi tessellation are convex. [Hint: Choose 2 points inside a polygon; show the line segment joining them is also inside.]

8. For a Poisson-Voronoi tessellation, condition on the event that a particle of the point process lies on the origin O. Consider the line process constructed by drawing the right-bisectors of all line-segments OC_i, where C_i $(i = 1, 2, \cdots)$ are the other particles listed in order of distance from O. Let $M(B_r)$ be the number of lines hitting the ball B_r. What is the distribution of $M(B_r)$? Define $R(t)$ as the distance from O in direction t until the first line is hit. Find the distribution of $R(t)$. Find the correlation coefficient of $R(t)$ and $R(t+h)$.

9. (Continuation) Let $\phi(t)$ be the angle between the ray emanating from O in direction t and the first line that the ray hits. Find the joint distribution of $R(t)$ and $\phi(t)$, and hence the marginal pdf of $\phi(t)$. [Note: Make any sensible definition of $\phi(t)$ when the ray hits a corner of the polygon, as it will with probability zero.]

10. Consider a rectangle D measuring 3 cms by 4 cms. A random transect of this rectangle is drawn. With probability one, it cuts precisely 2 of the rectangle's sides. Find the probabilities of the following three events, A, B and C.

$$\begin{aligned} &A : \text{one side cut is 3 cms, the other 4 cms} \\ &B : \text{both sides cut are 3 cms long} \\ &C : \text{both sides cut are 4 cms long.} \end{aligned}$$

11. H is a convex subset of a convex planar domain D. What is the probability that a random transect of D cuts H?

12. Two independent random transects of a convex planar domain D are made. What is the probability that the two lines cut each other within D?

13. We have proved in lectures that the mean length of a random transect drawn on a convex domain D is $\pi A/L$, where A and L are the area and perimeter of D. Given n independent random transects of D, let X be the number of intersection points of

these transects within D. Also define Y as the number of sub-regions of D formed by the n transects.

(a) Prove that, when $n = 3$, there is zero probability of the three transects crossing at a common point.

(b) Extend this result to prove that there is zero probability of *any* 3 transects crossing at a common point.

(c) Prove that, with probability 1, $Y = X + n + 1$. [*Hint:* Induction on n, using (b)?]

(d) Find $\mathrm{E}(X)$ and $\mathrm{E}(Y)$.

14. A 3–d convex body is projected onto a randomly-oriented planel by a parallel light beam to give a 2–d shape. There is a randomly-oriented line drawn on the plane. The 2–d shape is projected onto this line and the content C of the projected 1–d set observed. Show that the Cauchy-Kubota formula applied in two stages (from dimensions 3 to 2 then 2 to 1) gives the same result as when applied directly (dimension 3 to 1).

15. A white 3–d ball of radius r contains 2 black objects. One is a rectangular prism of sides a, b and c. The other is a cube of the same volume. An IUR random section of the sphere is made. What is the probability that it cuts (a) the rectangle (b) the cube. Let p be the probability that it cuts both (it is difficult to find p, so don't). In terms of p, a, b, c and r, find the probability that the cube is cut <u>given</u> that at least one of the objects is cut.

16. In 15, what is the proportion of the sphere taken up by objects. What is the expected proportion of black area on a random IUR section of the sphere?

17. D is a convex 2-d domain having area A and perimeter L. Other convex 2-d domains, W_1 and W_2, independently make IUR hits of D. The domain W_i has area A_i and perimeter L_i, $i = 1, 2$. What is the probability that $W_1 \cap W_2 \cap D \neq \phi$?

18. D is a $a \times b$ rectangle. A congruent copy of D, called D_1 say, makes an IUR hit of D. Let $N =$ the number of crossing points between the boundary ∂D and the boundary ∂D_1. Find $E(N)$.

19. A triangle ABC has angles α, β, γ and opposite sides a, b, c. This triangle is hit by an IUR transect $\mathcal{L}$.

(a) What is the probability that $\mathcal{L}$ hits the side AB?

(b) What is the probability that $\mathcal{L}$ hits both of the sides AB and AC?

(c) The hit by $\mathcal{L}$ partitions the triangle ABC into two parts, one being triangular. This triangular part contains just one of the original vertices. This could be either A, B or C but let us use the generic label V. Define the random variables X and Y as follows. $\mathcal{L}$ cuts the <u>left</u> 'arm' of V at P and the <u>right</u> 'arm' at Q. X is defined as the length of VP as a proportion of the left-arm's length. Y is the length of VQ as a proportion of the right-arm's length. Find $F(x, y)$, defined as $P\{X \leq x,\ Y \leq y\}$.