1 Unreliable Servers

In a single cluster of a Google competitor, there are a huge number of servers n, each with a uniform and independent probability of going down in a given day. On average, 4 servers go down in the cluster per day. As each cluster is responsible for a huge amount of internet traffic, it is fair to assume that n is a very large number. Recall that as $n \to \infty$, a $\text{Binom}(n, \lambda/n)$ distribution will tend towards a $\text{Poisson}(\lambda)$ distribution.

(a) What is an appropriate distribution to model the number of servers that crash on any given day for a certain cluster?

(b) Compute the expected value and variance of the number of crashed servers on a given day for a certain cluster.

(c) Compute the probability that fewer than 3 servers crashed on a given day for a certain cluster.

(d) Compute the probability at least 3 servers crashed on a given day for a certain cluster.

Solution:

(a) Because each server goes down independently of the other servers, and with the same probability, the number of servers that crash on a given day follows a binomial distribution $\text{Binom}(n, p)$, where n is the number of servers and p is the probability of each individual server crashing on any given day. Since on average, 4 servers crash per day, we have $p = \frac{4}{n}$. We are given that the number of servers in the cluster is large, so $n \gg p$ and we can model the number of servers that crash as a Poisson distribution with $\lambda = 4$.

(b) Recall that the expectation and variance of a Poisson distribution with parameter λ are both equal to λ and in this case $\lambda = 4$.

(c) To compute the probability that fewer than 3 servers went down, we must add the probabilities that 0 servers go down, 1 server goes down, and the probability that 2 servers go down. The PMF of the Poisson distribution is

$$\mathbb{P}[X = i] = \frac{\lambda^i}{i!} e^{-\lambda}.$$

Thus

$$\mathbb{P}[X = 0 \text{ or } X = 1 \text{ or } X = 2] = e^{-4} + 4e^{-4} + \frac{4^2}{2} e^{-4} = e^{-4} + 4e^{-4} + 8e^{-4} = 13e^{-4}.$$

(d) $1 - \mathbb{P}[\text{fewer than 3 servers crashed}] = 1 - 13e^{-4}$.

2 Class Enrollment

Lydia has just started her CalCentral enrollment appointment. She needs to register for a marine science class and CS 70. There are no waitlists, and she can attempt to enroll once per day in either class or both. The CalCentral enrollment system is strange and picky, so the probability of enrolling successfully in the marine science class on each attempt is μ and the probability of enrolling successfully in CS 70 on each attempt is λ. Also, these events are independent.

(a) Suppose Lydia begins by attempting to enroll in the marine science class everyday and gets enrolled in it on day M. What is the distribution of M?

(b) Suppose she is not enrolled in the marine science class after attempting each day for the first 5 days. What is the conditional distribution of M given $M > 5$?

(c) Once she is enrolled in the marine science class, she starts attempting to enroll in CS 70 from day $M+1$ and gets enrolled in it on day C. Find the expected number of days it takes Lydia to enroll in both the classes, i.e. $\mathbb{E}[C]$.

(d) Suppose instead of attempting one by one, Lydia decides to attempt enrolling in both the classes from day 1. Let M be the number of days it takes to enroll in the marine science class, and C be the number of days it takes to enroll in CS 70. What is the distribution of M and C now? Are they independent?

(e) Let X denote the day she gets enrolled in her first class and let Y denote the day she gets enrolled in both the classes. What is the distribution of X?

(f) What is the expected number of days it takes Lydia to enroll in both classes now, i.e. $\mathbb{E}[Y]$.

(g) What is the expected number of classes she will be enrolled in by the end of 14 days?

Solution:

(a) $M \sim \text{Geometric}(\mu)$.

(b) Given that $M > 5$, the random variable M takes values in $\{6, 7, \dots\}$. For $i = 6, 7, \dots,$

$$\mathbb{P}[M = i | M > 5] = \frac{\mathbb{P}[M = i \wedge M > 5]}{\mathbb{P}[M > 5]} = \frac{\mathbb{P}[M = i]}{\mathbb{P}[M > 5]} = \frac{\mu(1-\mu)^{i-1}}{(1-\mu)^5} = \mu(1-\mu)^{i-6}.$$

If K denotes the additional number of days it takes to get enrolled in the marine science class after day 5, i.e. $K = M - 5$, then conditioned on $M > 5$, the random variable K has the geometric distribution with parameter μ. Note that this is the same as the distribution of M. This is known as the memoryless property of geometric distribution.

(c) We have $C - M \sim \text{Geometric}(\lambda)$. Thus $\mathbb{E}[M] = 1/\mu$ and $\mathbb{E}[C-M] = 1/\lambda$. And hence $\mathbb{E}[C] = \mathbb{E}[M] + \mathbb{E}[C-M] = 1/\mu + 1/\lambda$.

(d) $M \sim \text{Geometric}(\mu)$, $C \sim \text{Geometric}(\lambda)$. Yes they are independent.

(e) We have $X = \min\{M,C\}$ and $Y = \max\{M,C\}$. We also use the following definition of the minimum:

$$\min(m,c) = \begin{cases} m & \text{if } m \leq c; \\ c & \text{if } m > c. \end{cases}$$

Now, for all $k \in \{1,2,\ldots\}$, $\min(M,C) = k$ is equivalent to $(M = k) \cap (C \geq k)$ or $(C = k) \cap (M > k)$. Hence,

$$\begin{aligned} \mathbb{P}[X = k] &= \mathbb{P}[\min(M,C) = k] \\ &= \mathbb{P}[(M = k) \cap (C \geq k)] + \mathbb{P}[(C = k) \cap (M > k)] \\ &= \mathbb{P}[M = k] \cdot \mathbb{P}[C \geq k] + \mathbb{P}[C = k] \cdot \mathbb{P}[M > k] \end{aligned}$$

(since M and C are independent)

$$= [(1-\mu)^{k-1}\mu](1-\lambda)^{k-1} + [(1-\lambda)^{k-1}\lambda](1-\mu)^{k}$$

(since M and C are geometric)

$$\begin{aligned} &= ((1-\mu)(1-\lambda))^{k-1}(\mu + \lambda(1-\mu)) \\ &= (1-\mu-\lambda+\lambda\mu)^{k-1}(\mu+\lambda-\mu\lambda). \end{aligned}$$

But this final expression is precisely the probability that a geometric r.v. with parameter $\mu + \lambda - \mu\lambda$ takes the value k. Hence $X \sim \text{Geom}(\mu + \lambda - \mu\lambda)$.

An alternative, slightly cleaner approach is to work with the *tail probabilities* of the geometric distribution, rather than with the usual point probabilities as above. In other words, we can work with $\mathbb{P}[X \geq k]$ rather than with $\mathbb{P}[X = k]$; clearly the values $\mathbb{P}[X \geq k]$ specify the values $\mathbb{P}[X = k]$ since $\mathbb{P}[X = k] = \mathbb{P}[X \geq k] - \mathbb{P}[X \geq (k+1)]$, so it suffices to calculate them instead. We then get the following argument:

$$\begin{aligned} \mathbb{P}[X \geq k] &= \mathbb{P}[\min(M,C) \geq k] \\ &= \mathbb{P}[(M \geq k) \cap (C \geq k)] \\ &= \mathbb{P}[M \geq k] \cdot \mathbb{P}[C \geq k] && \text{since } M,C \text{ are independent} \\ &= (1-\mu)^{k-1}(1-\lambda)^{k-1} && \text{since } M,C \text{ are geometric} \\ &= ((1-\mu)(1-\lambda))^{k-1} \\ &= (1-\mu-\lambda+\mu\lambda)^{k-1}. \end{aligned}$$

This is the tail probability of a geometric distribution with parameter $\mu + \lambda - \mu\lambda$, so we are done.

(f) From part (e) we get $\mathbb{E}[X] = 1/(\mu + \lambda - \mu\lambda)$. From part (d) we have $\mathbb{E}[M] = 1/\mu$ and $\mathbb{E}[C] = 1/\lambda$. We now observe that $\min\{m,c\} + \max\{m,c\} = m + c$. Using linearity of expectation we get $\mathbb{E}[X] + \mathbb{E}[Y] = \mathbb{E}[M] + \mathbb{E}[C]$. Thus $\mathbb{E}[Y] = 1/\mu + 1/\lambda - 1/(\mu + \lambda - \mu\lambda)$.

(g) Let I_M and I_C be the indicator random variables of the events "$M \leq 14$" and "$C \leq 14$" respectively. Then $I_M + I_C$ is the number of classes she will be enrolled in within 14 days. Hence the answer is $\mathbb{E}[I_M] + \mathbb{E}[I_C] = \mathbb{P}[M \leq 14] + \mathbb{P}[C \leq 14] = 1 - (1-\mu)^{14} + 1 - (1-\lambda)^{14}$

3 Short Answer

(a) Let X be uniform on the interval $[0,2]$, and define $Y = 2X + 1$. Find the PDF, CDF, expectation, and variance of Y.

(b) Let X and Y have joint distribution

$$f(x,y) = \begin{cases} cxy + 1/4 & x \in [1,2] \text{ and } y \in [0,2] \\ 0 & \text{else} \end{cases}$$

Find the constant c. Are X and Y independent?

(c) Let $X \sim \text{Exp}(3)$. What is the probability that $X \in [0,1]$? If I define a new random variable $Y = \lfloor X \rfloor$, for each $k \in \mathbb{N}$, what is the probability that $Y = k$? Do you recognize this (discrete) distribution?

(d) Let $X_i \sim \text{Exp}(\lambda_i)$ for $i = 1, \ldots, n$ be mutually independent. It is a (very nice) fact that $\min(X_1, \ldots, X_n) \sim \text{Exp}(\mu)$. Find μ.

Solution:

(a) Let's begin with the CDF. It will first be useful to recall that

$$F_X(t) = \mathbb{P}(X \leq t) = \begin{cases} 0 & t \leq 0 \\ \frac{t}{2} & t \in [0,2] \\ 1 & t \geq 1 \end{cases}.$$

Since Y is defined in terms of X, we can compute that

$$\begin{aligned} F_Y[t] = \mathbb{P}(Y \leq t) &= \mathbb{P}[2X + 1 \leq t] \\ &= \mathbb{P}\left[X \leq \frac{t-1}{2}\right] \\ &= F_X\left(\frac{t-1}{2}\right) \\ &= \begin{cases} 0 & t \leq 1 \\ \frac{t-1}{4} & t \in [1,5] \\ 1 & t \geq 5 \end{cases} \end{aligned}$$

where in the third line we have used the PDF for X. We know that the PDF can be found by taking the derivative of the CDF, so

$$f_Y(t) = \frac{dF_Y(t)}{dt} = \begin{cases} \frac{1}{4} & t \in [1,5] \\ 0 & \text{else} \end{cases}. \tag{1}$$

By linearity of expectation $\mathbb{E}[Y] = \mathbb{E}[2X + 1] = 2\,\mathbb{E}[X] + 1 = 3$, and similarly

$$\text{Var}(Y) = \text{Var}(2X + 1) = 4\,\text{Var}(X) = 4\frac{4}{12} = \frac{4}{3}$$

(b) To find the correct constant, we use the fact that a PDF must integrate to one. In particular,

$$1 = \int_1^2 \int_0^2 (cxy + 1/4)\,dy\,dx = 3c + 1/2,$$

so $c = 1/6$. In order to check independence, we need to first find the marginal distributions of X and Y:

$$f_X(x) = \int_0^2 f(x,y)\,dy = 1/2 + x/3$$
$$f_Y(y) = \int_1^2 f(x,y)\,dx = 1/4 + y/4.$$

Since $f_X(x)f_Y(y) = 1/8 + y/8 + x/12 + xy/12 \neq 1/4 + xy/6 = f(x,y)$, the random variables are not independent.

(c) Since $X \sim \text{Exp}(3)$, the PDF of X is $f(x) = 3e^{-3x}$. Thus we have

$$\mathbb{P}[X \in [0,1]] = \int_0^1 f(x)\,dx = \int_0^1 3e^{-3x} = -e^{-3x}\big|_0^1 = 1 - e^{-3}.$$

Similarly, if $Y = \lfloor X \rfloor$, then $Y = k$ exactly when $X \in [k, k+1)$, so

$$\begin{aligned}
\mathbb{P}[Y = k] &= \mathbb{P}[X \in [k, k+1)] \\
&= \int_k^{k+1} f(x)\,dx \\
&= -e^{-3x}\big|_k^{k+1} \\
&= e^{-3(k)} - e^{-3(k+1)} \\
&= e^{-3k}\left(1 - e^{-3}\right) = (e^{-3})^k\left(1 - e^{-3}\right).
\end{aligned}$$

In other words, $Y \sim \text{Geometric}(1 - e^{-3}) - 1$.

(d) Since the X_i are independent,

$$\begin{aligned}
&\mathbb{P}[\min(X_1, \ldots, X_n) \leq t] \\
&\quad = 1 - \mathbb{P}[X_1 > t, X_2 > t, \cdots X_n > t] \\
&\quad = 1 - \mathbb{P}[X_1 > t]\mathbb{P}[X_2 > t]\cdots\mathbb{P}[X_n > t] && \text{independence} \\
&\quad = 1 - e^{-\lambda_1 t}e^{-\lambda_2 t}\cdots e^{-\lambda_n t} \\
&\quad = 1 - e^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)t}.
\end{aligned}$$

This is exactly the CDF of an $\text{Exp}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$ random variable, so $\mu = \lambda_1 + \cdots + \lambda_n$.

4 Useful Uniforms

Let X be a continuous random variable whose image is all of $\mathbb{R}$; that is, $\mathbb{P}[X \in (a,b)] > 0$ for all $a,b \in \mathbb{R}$ and $a \neq b$.

(a) Give an example of a distribution that X could have, and one that it could not.

(b) Show that the CDF F of X is strictly increasing. That is, $F(x+\varepsilon) > F(x)$ for any $\varepsilon > 0$. Argue why this implies that $F : \mathbb{R} \to (0,1)$ must be invertible.

(c) Let U be a uniform random variable on $(0,1)$. What is the distribution of $F^{-1}(U)$?

(d) Your work in part (c) shows that in order to sample X, it is enough to be able to sample U. If X was a discrete random variable instead, taking finitely many values, can we still use U to sample X?

Solution:

(a) Any random variable with density $f(x) > 0$ for all x works as a positive example; e.g. $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ (corresponding to the normal distribution) or $f(x) = \begin{cases} 1/2, & \text{if } |x| < 1, \\ \frac{1}{4|x|^2}, & \text{if } |x| \geq 1 \end{cases}$.

Any distribution of density f such that $f(x) = 0$ for all $x \in (a,b)$ for some $a,b \in \mathbb{R}, a \neq b$ works as a negative example; e.g. $f(x) = \begin{cases} e^{-x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise} \end{cases}$ (corresponding to an exponential random variable) or $f(x) = \begin{cases} 1, & \text{if } x \in [0,1], \\ 0, & \text{otherwise} \end{cases}$ (corresponding to a uniform variable on $[0,1]$).

(b) $F(x+\varepsilon) = \mathbb{P}[X \leq x+\varepsilon] = \mathbb{P}[X \leq x] + \mathbb{P}[X \in (x, x+\varepsilon]] \geq F(x) + \mathbb{P}[X \in (x, x+\varepsilon)] > F(x)$, where in the very last inequality we used the fact that $\mathbb{P}[X \in (a,b)] > 0$ with $a = x$ and $b = x+\varepsilon$. To show invertibility, we need to show (i) injectivity and (ii) surjectivity. (i): If $x \neq y$, then either $x < y$ or $y < x$ and so either $F(x) < F(y)$ or $F(y) < F(x)$. In either case, $F(x) \neq F(y)$, and so F must be injective. (ii) F is continuous (in fact, differentiable with derivative f), approaching 1 as $x \to \infty$, and approaching 0 as $x \to -\infty$. Therefore, it must assume all values between 0 and 1, and hence is surjective.

(c) $\mathbb{P}\left[F^{-1}(U) \leq x\right] = \mathbb{P}[U \leq F(x)] = F(x)$, where $\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$ since F is strictly increasing. Thus $F^{-1}(U)$ and X have the very same CDF, which means that $F^{-1}(U)$ and X share the same distribution.

(d) Yes, we can! Assume X took values in a discrete set $\mathscr{A} = \{a_1, a_2, \ldots, a_n\} \subset \mathbb{R}$ with probabilities $\mathbb{P}[X = a_k] = p_k$. Then mimicking the argument from part (c), we can define $G : [0,1] \to \mathscr{A}$

as

$$G(x) = \begin{cases} a_1, & \text{if } x \le p_1, \\ a_2, & \text{if } x \in (p_1, p_1 + p_2], \\ a_3, & \text{if } x \in (p_1 + p_2, p_1 + p_2 + p_3], \\ \vdots & \vdots \\ a_{n-1}, & \text{if } x \in (\sum_{k=1}^{n-2} p_k, \sum_{k=1}^{n-1} p_k], \\ a_n, & \text{if } x \in (\sum_{k=1}^{n-1} p_k, 1] \end{cases}$$

(draw a picture of G's graph!), for which we have $\mathbb{P}[G(U) = a_k] = \sum_{j=1}^{k} p_j - \sum_{j=1}^{k-1} p_j = p_k = \mathbb{P}[X = a_k]$. That is, $G(U)$ and X have the same distribution as desired.

5 It's Raining Fish

A hurricane just blew across the coast and flung a school of fish onto the road nearby the beach. The road starts at your house and is infinitely long. We will label a point on the road by its distance from your house (in miles). For each $n \in \mathbb{N}$, the number of fish that land on the segment of the road $[n, n+1]$ is independently $pois(\lambda)$ and each fish that is flung into that segment of the road lands uniformly at random within the segment. Keep in mind that you can cite any result from lecture or discussion without proof.

(a) What is the distribution of the number of fish arriving in segment $[0, n]$ of the road, for some $n \in \mathbb{N}$.

(b) Let $[a, b]$ be an interval in $[0, 1]$. What is the distribution of the number of fish that lands in the segment $[a, b]$ of the road?

(c) Let $[a, b]$ be any interval such that $a \ge 0$. What is the distribution of the number of fish that land in $[a, b]$?

(d) Suppose you take a stroll down the road. What is the distribution of the distance you walk (in miles) until you encounter the first fish? Justify with proof.

(e) Suppose you encounter a fish at distance x. What is the distribution of the distance you walk until you encounter the next fish?

Solution:

(a) From lecture, we learned that if X and Y are independent, and $X \sim Pois(\lambda)$ and $Y \sim Pois(\mu)$, then $X + Y \sim Pois(\lambda + \mu)$. We know the number of fish to land in the segment $[0, n]$ is the sum of the number of fish to land in $[i, i+1]$ for each $i \in [0, 1, \ldots n-1]$. Thus the number of fish is in $[0, n]$ is $Pois(n\lambda)$

(b) The probability that a particular fish lands in the interval $[a, b]$ is $b - a$ since it's location is uniformly distributed within $[0, 1]$. Thus, the distribution is $Pois((b-a)\lambda)$

(c) The answer is still $Pois((b-a)\lambda)$. Clearly, this is true if $[a,b]$ is contained within some interval $[n,n+1]$. If it's not, then let i be the smallest integer such that $i \geq a$ and let j be the largest integer such that such that $j \leq b$. Then the distribution is

$$Pois((i-a)\lambda) + Pois((j-i)\lambda) + Pois((b-j)\lambda) = Pois((b-a)\lambda)$$

(d) The distance is $expo(\lambda)$. To prove this, it suffices to show that the cdf matches the exponential cdf. Let X be the distance of the first fish from the house. Note that $\mathbb{P}[X \geq t] = \mathbb{P}[\text{no fish in } [0,t]]$. By the previous parts, we know that the number of fish in $[0,t]$ is $Pois(\lambda t)$, which is equaled to 0 with probability

$$\frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

Thus, we have that $Pr[X < t] = 1 - e^{-\lambda t}$ which is exactly the exponential cdf.

(e) Still $expo(\lambda)$. Using the same logic as the previous part, we have that

$$\mathbb{P}[X \geq t]$$

is equaled to the probability that no fish land in the segment $[x, x+t]$ which is equaled to $e^{-\lambda t}$ because the number of fish in that segment is $Pois(-\lambda t)$.

6 Exponential Expectation

(a) Let $X \sim \text{Exp}(\lambda)$. Use induction to show that $\mathbb{E}[X^k] = k!/\lambda^k$ for every $k \in \mathbb{N}$.

(b) For any $|t| < \lambda$, compute $\mathbb{E}[e^{tX}]$ directly from the definition of expectation.

(c) Using part (a), compute $\sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} t^k$.

(d) Let $M(t) = \mathbb{E}[e^{tX}]$ be a function defined for all t such that $|t| < \lambda$. What is $\left.\frac{dM(t)}{dt}\right|_{t=0}$? What is $\left.\frac{d^2M(t)}{dt^2}\right|_{t=0}$? How does each of these relate to the mean and variance of an $\text{Exp}(\lambda)$ distribution?

Solution:

(a) The base case is $\mathbb{E}[X] = 1/\lambda$, which we already know. Using integration by parts,

$$\begin{aligned}
\mathbb{E}[X^{k+1}] &= \int_0^\infty x^{k+1} \cdot \lambda e^{-\lambda x}\, \mathrm{d}x \\
&= -x^{k+1} e^{-\lambda x}\Big|_0^\infty + (k+1) \int_0^\infty x^k e^{-\lambda x}\, \mathrm{d}x \\
&= \frac{k+1}{\lambda} \int_0^\infty x^k \cdot \lambda e^{-\lambda x}\, \mathrm{d}x \\
&= \frac{k+1}{\lambda} \mathbb{E}[X^k] \\
&= \frac{(k+1)!}{\lambda^{k+1}}
\end{aligned}$$

which proves the inductive step.

(b) For any $|t| < \lambda$.

$$\begin{aligned}\mathbb{E}[\exp(tX)] &= \int_0^\infty e^{tx}\lambda e^{-\lambda x}\,\mathrm{d}x \\ &= \frac{\lambda}{\lambda - t}\int_0^\infty (\lambda - t)e^{-(\lambda - t)x}\,\mathrm{d}x \\ &= \frac{\lambda}{\lambda - t} \\ &= \frac{1}{1 - t/\lambda}\end{aligned}$$

(c) We have,

$$\sum_{k=0}^\infty \frac{\mathbb{E}[X^k]}{k!}t^k = \sum_{k=0}^\infty \frac{t^k}{\lambda^k} = \frac{1}{1 - t/\lambda}$$

for any $|t| < \lambda$ (if $|t| \geq \lambda$ then this series does not converge). This is the same as what we found in part (b)! Recall the power series expansion

$$e^x = \sum_{k=0}^\infty \frac{x^k}{k!}.$$

If X is any random variable, and we plug in tX for x in this identity and take expectations (remembering linearity of course!), we get

$$\mathbb{E}[\exp(tX)] = \mathbb{E}\left[\sum_{k=0}^\infty \frac{t^k X^k}{k!}\right] = \sum_{k=0}^\infty \frac{t^k\,\mathbb{E}[X^k]}{k!}$$

for whichever t the series on the right side converges.

(d)

$$\left.\frac{dM(t)}{dt}\right|_{t=0} = \left.\frac{\lambda}{(\lambda - t)^2}\right|_{t=0} = \frac{1}{\lambda} = \mu_1$$

$$\left.\frac{d^2M(t)}{dt^2}\right|_{t=0} = \left.\frac{d}{dt}\frac{\lambda}{(\lambda - t)^2}\right|_{t=0} = \left.\frac{2\lambda}{(\lambda - t)^3}\right|_{t=0} = \frac{2}{\lambda^2} = \mu_2$$

μ_1 is the mean of an $\mathrm{Exp}(\lambda)$ distribution, and $\mu_2 - \mu_1^2$ is the variance of that distribution.

μ_2 is called the second moment of the distribution. In general, $\left.\frac{d^nM(t)}{dt^n}\right|_{t=0} = \mu_n = \mathbb{E}[X^n]$ is called the n^{th} moment of the random variable X, and $M(t) = \mathbb{E}[e^{tX}]$ is called the moment-generating function (mgf) of X. Just like the pdf and cdf, the mgf of a distribution also uniquely characterizes the probability distribution.

7 Noisy Love

Suppose you have confessed to your love interest on Valentine's Day and you are waiting to hear back. Your love interest is trying to send you a binary message: "0" means that your love interest is not interested in you, while "1" means that your love interest reciprocates your feelings. Let X be your love interest's message for you. Your current best guess of X has $\mathbb{P}(X = 0) = 0.7$ and $\mathbb{P}(X = 1) = 0.3$. Unfortunately, your love interest sends you the message through a noisy channel, and instead of receiving the message X, you receive the message $Y = X + \varepsilon$, where ε is independent Gaussian noise with mean 0 and variance 0.49.

(a) First, you decide upon the following rule: if you observe $Y > 0.5$, then you will assume that your love interest loves you back, whereas if you observe $Y \le 0.5$, then you will assume that your love interest is not interested in you. What is the probability that you are correct using this rule? (Express your answer in terms of the CDF of the standard Gaussian distribution $\Phi(z) = \mathbb{P}(\mathcal{N}(0,1) \le z)$, and then evaluate your answer numerically.)

(b) Suppose you observe $Y = 0.6$. What is the probability that your love interest loves you back? [*Hint*: This problem requires conditioning on an event of probability 0, namely, the event $\{Y = 0.6\}$. To tackle this problem, think about conditioning on the event $\{Y \in [0.6, 0.6+\delta]\}$, where $\delta > 0$ is small, so that $f_Y(0.6) \cdot \delta \approx \mathbb{P}(Y \in [0.6, 0.6+\delta])$, and then apply Bayes Rule.]

(c) Suppose you observe $Y = y$. For what values is it more likely than not that your love interest loves you back? [*Hint*: As before, instead of considering $\{Y = y\}$, you can consider the event $\{Y \in [y, y+\delta]\}$ for small $\delta > 0$. So, when is $\mathbb{P}(X = 1 \mid Y \in [y, y+\delta]) \ge \mathbb{P}(X = 0 \mid Y \in [y, y+\delta])$?]

(d) Your new rule is to assume that your love interest loves you back if (based on the value of Y that you observe) it is more likely than not that your love interest loves you back. Under this new rule, what is the probability that you are correct?

Solution:

(a) The probability that you are correct is

$$\begin{aligned}
&\mathbb{P}(X = 0)\mathbb{P}(Y \le 0.5 \mid X = 0) + \mathbb{P}(X = 1)\mathbb{P}(Y > 0.5 \mid X = 1) \\
&\quad = 0.7\mathbb{P}(\mathcal{N}(0, 0.49) \le 0.5) + 0.3\mathbb{P}(\mathcal{N}(1, 0.49) > 0.5) \\
&\quad = 0.7\mathbb{P}\left(\mathcal{N}(0,1) \le \frac{0.5}{0.7}\right) + 0.3\mathbb{P}\left(\mathcal{N}(0,1) > -\frac{0.5}{0.7}\right) = \mathbb{P}\left(\mathcal{N}(0,1) \le \frac{5}{7}\right) = \Phi\left(\frac{5}{7}\right) \\
&\quad \approx 0.762.
\end{aligned}$$

(b) By conditioning on $\{Y \in [0.6, 0.6+\delta]\}$, we have

$$
\begin{aligned}
&\mathbb{P}(X = 1 \mid Y \in [0.6, 0.6+\delta]) \\
&\quad = \frac{\mathbb{P}(X=1)\mathbb{P}(Y \in [0.6,0.6+\delta] \mid X=1)}{\mathbb{P}(X=0)\mathbb{P}(Y \in [0.6,0.6+\delta] \mid X=0) + \mathbb{P}(Y=1)\mathbb{P}(Y \in [0.6,0.6+\delta] \mid X=1)} \\
&\quad = \frac{\mathbb{P}(X=1) f_{Y|1}(0.6)\delta}{\mathbb{P}(X=0) f_{Y|0}(0.6)\delta + \mathbb{P}(X=1) f_{Y|1}(0.6)\delta} \\
&\quad = \frac{\mathbb{P}(X=1) f_{Y|1}(0.6)}{\mathbb{P}(X=0) f_{Y|0}(0.6) + \mathbb{P}(X=1) f_{Y|1}(0.6)},
\end{aligned}
$$

where $f_{Y|0}$ is the density of a Gaussian with mean 0 and variance 0.49, and $f_{Y|1}$ is the density of a Gaussian with mean 1 and variance 0.49. Although the expression above may look intimidating, *this is just Bayes rule where $\mathbb{P}(Y = 0.6 \mid X = x)$ has been replaced with* $f_{Y|x}(0.6)$. The moral of the story is that conditioning in continuous probability seems strange at first, but it is essentially the same as conditioning in discrete probability, with densities taking the place of probability mass functions.

Now, filling in the probabilities, we have

$$
\begin{aligned}
&\mathbb{P}(X = 1 \mid Y = 0.6) \\
&\quad = \frac{0.3 \cdot (2\pi)^{-1/2} \exp(-0.4^2/(2 \cdot 0.49))}{0.7 \cdot (2\pi)^{-1/2} \exp(-0.6^2/(2 \cdot 0.49)) + 0.3 \cdot (2\pi)^{-1/2} \exp(-0.4^2/(2 \cdot 0.49))} \approx 0.345.
\end{aligned}
$$

See what happened here? Before, you thought $\mathbb{P}(X = 1) = 0.3$. Observing $Y = 0.6$ gives you slightly more evidence in favor of your love interest loving you back, which increases your belief to $\mathbb{P}(X = 1 \mid Y = 0.6) = 0.345$.

(c) We are looking for

$$\mathbb{P}(X = 1 \mid Y \in [y, y+\delta]) \geq \mathbb{P}(X = 0 \mid Y \in [y, y+\delta])$$

which is equivalent to

$$\mathbb{P}(X = 1 \mid Y \in [y, y+\delta]) \geq \frac{1}{2}.$$

Now, we can compute the LHS as in the previous part:

$$
\begin{aligned}
\mathbb{P}(X = 1 \mid Y \in [y, y+\delta]) &= \frac{\mathbb{P}(X=1) f_{Y|1}(y)}{\mathbb{P}(X=0) f_{Y|0}(y) + \mathbb{P}(X=1) f_{Y|1}(y)} \\
&= \frac{0.3 \exp(-(1-y)^2/0.98)}{0.7 \exp(-y^2/0.98) + 0.3 \exp(-(1-y)^2/0.98)} \\
&= \frac{1}{1 + (0.7/0.3) \exp(((1-y)^2 - y^2)/0.98)}.
\end{aligned}
$$

In order to make the RHS $\geq 1/2$, we need:

$$\begin{aligned}
\frac{0.7}{0.3}\exp\left(\frac{(1-y)^2-y^2}{0.98}\right) &\leq 1 \\
\exp\left(\frac{(1-y)^2-y^2}{0.98}\right) &\leq \frac{3}{7} \\
\frac{(1-y)^2-y^2}{0.98} &\leq \ln\frac{3}{7} \\
1-2y &\leq 0.98\ln\frac{3}{7}
\end{aligned}$$

which gives the condition

$$y \geq \frac{1}{2}\left(1-0.98\ln\frac{3}{7}\right) \approx 0.915.$$

So, the new rule is to assume that your love interest loves you back if and only if you observe a message which is ≥ 0.915.

(d) As in the first part,

$$\begin{aligned}
&\mathbb{P}(X=0)\mathbb{P}(Y \leq 0.915 \mid X=0)+\mathbb{P}(X=1)\mathbb{P}(Y > 0.915 \mid X=1) \\
&= 0.7\mathbb{P}(\mathcal{N}(0,0.49) \leq 0.915)+0.3\mathbb{P}(\mathcal{N}(1,0.49) > 0.915) \\
&= 0.7\mathbb{P}\left(\mathcal{N}(0,1) \leq \frac{0.915}{0.7}\right)+0.3\mathbb{P}\left(\mathcal{N}(0,1) > -\frac{0.085}{0.7}\right) \\
&= 0.7\Phi\left(\frac{0.915}{0.7}\right)+0.3\Phi\left(\frac{0.085}{0.7}\right) \approx 0.798.
\end{aligned}$$

As you can see, this strategy performs better than the first part.

8 Sum of Independent Gaussians

In this question, we will introduce an important property of the Gaussian distribution: the sum of independent Gaussians is also a Gaussian.

Let X and Y be independent standard Gaussian random variables. Recall that the density of the standard Gaussian is

$$f(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right).$$

(a) What is the joint density of X and Y?

(b) Observe that the joint density of X and Y, $f_{X,Y}(x,y)$, only depends on the quantity x^2+y^2, which is the distance from the origin. In other words, the Gaussian is *rotationally symmetric*. Next, we will try to find the density of $X+Y$. To do this, draw a picture of the Cartesian plane and draw the region $x+y \leq c$, where c is a real number of your choice.

(c) Now, rotate your picture clockwise by $\pi/4$ so that the line $X+Y=c$ is now vertical. Redraw your figure. Let X' and Y' denote the random variables which correspond to the $\pi/4$ clockwise rotation of (X,Y) and express the new shaded region in terms of X' and Y'.

(d) By rotational symmetry of the Gaussian, (X',Y') has the same distribution as (X,Y). Argue that $X+Y$ has the same distribution as $\sqrt{2}Z$, where Z is a standard Gaussian. This proves the following important fact: *the sum of independent Gaussians is also a Gaussian*. Notice that $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$ and $X+Y \sim \mathcal{N}(0,2)$. *In general, if X and Y are independent Gaussians, then $X+Y$ is a Gaussian with mean $\mu_X+\mu_Y$ and variance $\sigma_X^2+\sigma_Y^2$.*

(e) Recall the CLT:

If $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 < \infty$, then:

$$\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{in distribution}} \mathcal{N}(0,1) \qquad \text{as } n\to\infty.$$

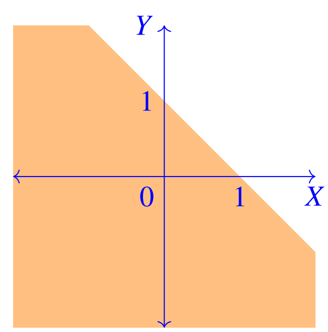
Prove that the CLT holds for the special case when the X_i are i.i.d. $\mathcal{N}(0,1)$.

Solution:

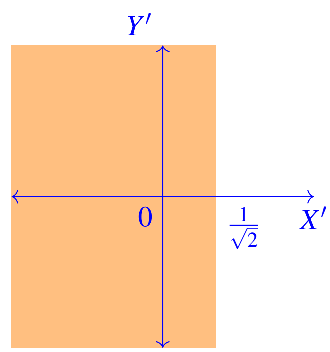
(a) By independence, we have

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right).$$

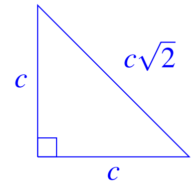
(b) We draw the line for $c=1$.



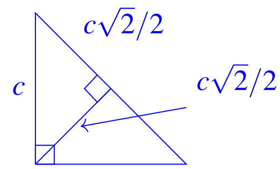
(c) Here is the new figure after the rotation (for $c=1$).



For general $c \in \mathbb{R}$, the new region is $\{X' \le c/\sqrt{2}\}$. To see why, draw the triangle: We want



to find the distance between the origin and the long side of the triangle, and we can do so by adding a diagonal:



(d) We observe that $\mathbb{P}(X+Y \le c) = \mathbb{P}(X' \le c/\sqrt{2}) = \mathbb{P}(\sqrt{2}X' \le c)$, where X' is a standard Gaussian by rotational symmetry, so this proves the claim.

(e) Here, $\mu = 0$ and $\sigma = 1$. So, by the previous part,

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \sim \frac{1}{\sqrt{n}}\mathcal{N}(0,n) \sim \mathcal{N}(0,1).$$

9 Homework Process and Study Group

You must describe your homework process and study group in order to receive credit for this question.