Contents

3.31 Theorem $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e.$

Proof Let

$$s_n = \sum_{k=0}^{n}\frac{1}{k!}, \qquad t_n = \left(1+\frac{1}{n}\right)^n.$$

By the binomial theorem,

$$t_n = 1 + 1 + \frac{1}{2!}\left(1-\frac{1}{n}\right) + \frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right) + \cdots$$
$$+ \frac{1}{n!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{n-1}{n}\right).$$

Hence $t_n \le s_n$, so that

(14) $$\limsup_{n\to\infty} t_n \le e,$$

by Theorem 3.19. Next, if $n \ge m$,

$$t_n \ge 1 + 1 + \frac{1}{2!}\left(1-\frac{1}{n}\right) + \cdots + \frac{1}{m!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{m-1}{n}\right).$$

Let $n \to \infty$, keeping m fixed. We get

$$\liminf_{n\to\infty} t_n \ge 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{m!},$$

so that

$$s_m \le \liminf_{n\to\infty} t_n.$$

Letting $m \to \infty$, we finally get

(15) $$e \le \liminf_{n\to\infty} t_n.$$

The theorem follows from (14) and (15).

3.37 Theorem *For any sequence $\{c_n\}$ of positive numbers,*

$$\liminf_{n\to\infty} \frac{c_{n+1}}{c_n} \le \liminf_{n\to\infty} \sqrt[n]{c_n},$$

$$\limsup_{n\to\infty} \sqrt[n]{c_n} \le \limsup_{n\to\infty} \frac{c_{n+1}}{c_n}.$$

Proof We shall prove the second inequality; the proof of the first is quite similar. Put

$$\alpha = \limsup_{n\to\infty} \frac{c_{n+1}}{c_n}.$$

If $\alpha = +\infty$, there is nothing to prove. If α is finite, choose $\beta > \alpha$. There is an integer N such that

$$\frac{c_{n+1}}{c_n} \le \beta$$

for $n \ge N$. In particular, for any $p > 0$,

$$c_{N+k+1} \le \beta c_{N+k} \qquad (k = 0, 1, \ldots, p-1).$$

Multiplying these inequalities, we obtain

$$c_{N+p} \le \beta^p c_N,$$

or

$$c_n \le c_N \beta^{-N} \cdot \beta^n \qquad (n \ge N).$$

Hence

$$\sqrt[n]{c_n} \le \sqrt[n]{c_N \beta^{-N}} \cdot \beta,$$

so that

(18) $$\limsup_{n\to\infty} \sqrt[n]{c_n} \le \beta,$$

by Theorem 3.20(*b*). Since (18) is true for every $\beta > \alpha$, we have

$$\limsup_{n\to\infty} \sqrt[n]{c_n} \le \alpha.$$

4. A crime is committed by one of two suspects, A and B. Initially, there is equal evidence against both of them. In further investigation at the crime scene, it is found that the guilty party had a blood type found in 10% of the population. Suspect A does match this blood type, whereas the blood type of Suspect B is unknown.

   (a) Given this new information, what is the probability that A is the guilty party?

   Let M be the event that A's blood type matches the guilty party's and for brevity, write A for "A is guilty" and B for "B is guilty". By Bayes' Rule,

$$P(A|M) = \frac{P(M|A)P(A)}{P(M|A)P(A) + P(M|B)P(B)} = \frac{1/2}{1/2 + (1/10)(1/2)} = \frac{10}{11}.$$

   (We have $P(M|B) = 1/10$ since, given that B is guilty, the probability that A's blood type matches the guilty party's is the same probability as for the general population.)

M be the event that A's blood type matches the guilty party's event1/10

Let U_1, U_2, U_3 be i.i.d. Unif$(0, 1)$, and let $L = \min(U_1, U_2, U_3)$, $M = \max(U_1, U_2, U_3)$

(a) Find the marginal CDF and marginal PDF of M, and the joint CDF and joint PDF of L, M.

Hint: for the latter, start by considering $P(L \geq l, M \leq m)$.

The event $M \leq m$ is the same as the event that all 3 of the U_j are at most m, so the CDF of M is $F_M(m) = m^3$ and the PDF is $f_M(m) = 3m^2$, for $0 \leq m \leq 1$.

The event $L \geq l, M \leq m$ is the same as the event that all 3 of the U_j are between l and m (inclusive), so

$$P(L \geq l, M \leq m) = (m - l)^3$$

for $m \geq l$ with $m, l \in [0, 1]$. By the axioms of probability, we have

$$P(M \leq m) = P(L \leq l, M \leq m) + P(L > l, M \leq m).$$

So the joint CDF is

$$P(L \leq l, M \leq m) = m^3 - (m - l)^3,$$

for $m \geq l$ with $m, l \in [0, 1]$. The joint PDF is obtained by differentiating this with respect to l and then with respect to m (or vice versa):

$$f(l, m) = 6(m - l),$$

for $m \geq l$ with $m, l \in [0, 1]$. As a check, note that getting the marginal PDF of M by finding $\int_0^m f(l, m)dl$ does recover the PDF of M (the limits of integration are from 0 to m since the min can't be more than the max).

6. Emails arrive in an inbox according to a Poisson process with rate λ (so the number of emails in a time interval of length t is distributed as $\text{Pois}(\lambda t)$, and the numbers of emails arriving in disjoint time intervals are independent). Let X, Y, Z be the numbers of emails that arrive from 9 am to noon, noon to 6 pm, and 6 pm to midnight (respectively) on a certain day.

(a) Find the joint PMF of X, Y, Z.

Since $X \sim \text{Pois}(3\lambda), Y \sim \text{Pois}(6\lambda), Z \sim \text{Pois}(6\lambda)$ independently, the joint PMF is

$$P(X = x, Y = y, Z = z) = \frac{e^{-3\lambda}(3\lambda)^x}{x!}\frac{e^{-6\lambda}(6\lambda)^y}{y!}\frac{e^{-6\lambda}(6\lambda)^z}{z!},$$

for any nonnegative integers x, y, z.

(b) Find the conditional joint PMF of X, Y, Z given that $X + Y + Z = 36$.

Let $T = X + Y + Z \sim \text{Pois}(15\lambda)$, and suppose that we observe $T = t$. The conditional PMF is 0 for $x + y + z \neq t$. For $x + y + z = t$,

$$\begin{aligned} P(X = x, Y = y, Z = z|T = t) &= \frac{P(T = t|X = x, Y = y, Z = z)P(X = x, Y = y, Z}{P(T = t)} \\ &= \frac{\frac{e^{-3\lambda}(3\lambda)^x}{x!}\frac{e^{-6\lambda}(6\lambda)^y}{y}\frac{e^{-6\lambda}(6\lambda)^z}{z!}}{\frac{e^{-15\lambda}(15\lambda)^t}{t!}} \\ &= \frac{t!}{x!y!z!}\left(\frac{3}{15}\right)^x\left(\frac{6}{15}\right)^y\left(\frac{6}{15}\right)^z. \end{aligned}$$

Thus, (X, Y, Z) is conditionally Multinomial given $T = t$, and we have that (X, Y, Z) is conditionally $\text{Multinomial}(36, (\frac{1}{5}, \frac{2}{5}, \frac{2}{5}))$ given $T = 36$.

(c) Find the conditional PMF of $X + Y$ given that $X + Y + Z = 36$, and find $E(X+Y|X+Y+Z = 36)$ and $\text{Var}(X+Y|X+Y+Z = 36)$ (conditional expectation

and conditional variance given an event are defined in the same way as expectation and variance, using the conditional distribution given the event in place of the unconditional distribution).

Let $W = X + Y$ and $T = X + Y + Z$. Using the story of the Multinomial and Part (b), we can merge the categories "9 am to noon" and "noon to 6 pm" to get

$$W|T = 36 \sim \text{Bin}\left(36, \frac{9}{15}\right).$$

Therefore, $E(W|T = 36) = 36 \cdot \frac{9}{15} = 21.6$ and $\text{Var}(W|T = 36) = 36 \cdot \frac{9}{15} \cdot \frac{6}{15} = 8.64$.

Prob. 19, Sec. 6.1, Pg. 295: A 3 by 3 matrix B is known to have eigenvalues $0, 1, 2$. This is information is enough to find three of these (give the answers where possible):

(a) the rank of B,

(b) the determinant of $B^{\mathrm{T}}B$,

(c) the eigenvalues of $B^{\mathrm{T}}B$,

(d) the eigenvalues of $(B^2 + I)^{-1}$.

$\boxed{\text{Solution}}$ (4 pts.): (a) The rank is at most 2 since B is singular as 0 is an eigenvalue. The rank is not 0 since B is not 0 as B has a nonzero eigenvalue. The rank is not 1 since a rank-1 matrix has only one nonzero eigenvalue as every eigenvector lies in the column space. Thus the rank is 2.

(b) We have $\det B^{\mathrm{T}}B = \det B^{\mathrm{T}} \det B = (\det B)^2 = 0 \cdot 1 \cdot 2 = 0$.

(c) There is not enough information to find the eigenvalues of $B^{\mathrm{T}}B$. For example,

$$\text{if } B = \begin{bmatrix} 0 & & \\ & 1 & \\ & & 2 \end{bmatrix}, \text{ then } B^{\mathrm{T}}B = \begin{bmatrix} 0 & & \\ & 1 & \\ & & 4 \end{bmatrix}; \quad \text{if } B = \begin{bmatrix} 0 & 1 & \\ & 1 & \\ & & 2 \end{bmatrix}, \text{ then } B^{\mathrm{T}}B = \begin{bmatrix} 0 & & \\ & 2 & \\ & & 4 \end{bmatrix}.$$

However, the eigenvalues of a triangular matrix are its diagonal entries.

(d) If $Ax = \lambda x$, then $x = \lambda A^{-1}x$; also, any polynomial $p(t)$ yields $p(A)x = p(\lambda)x$. Hence the eigenvalues of $(B^2 + I)^{-1}$ are $1/(0^2 + 1)$ and $1/(1^2 + 1)$ and $1/(2^2 + 1)$, or 1 and 1/2 and 1/5.

Every system with no solution is like the one in Problem 6. There are numbers $y_1, \ldots, y_m$ that multiply the m equations so they add up to $0 = 1$. This is called **Fredholm's Alternative**:

Exactly one of these problems has a solution

$$\boldsymbol{Ax = b} \quad \textbf{OR} \quad \boldsymbol{A^{\mathrm{T}}y = 0} \text{ with } \boldsymbol{y^{\mathrm{T}}b} = 1.$$

If $\boldsymbol{b}$ is not in the column space of A, it is not orthogonal to the nullspace of A^{T}. Multiply the equations $x_1 - x_2 = 1$ and $x_2 - x_3 = 1$ and $x_1 - x_3 = 1$ by numbers y_1, y_2, y_3 chosen so that the equations add up to $0 = 1$.

A total of 20 bags of Haribo gummi bears are randomly distributed to the 20 students in a certain Stat 110 section. Each bag is obtained by a random student, and the outcomes of who gets which bag are independent. Find the average number of bags of gummi bears that the first three students get in total, and find the average number of students who get at least one bag.

Let X_j be the number of bags of gummi bears that the jth student gets, and let I_j be the indicator of $X_j \geq 1$. Then $X_j \sim \text{Bin}(20, \frac{1}{20})$, so $E(X_j) = 1$. So $E(X_1 + X_2 + X_3) = 3$ by linearity.

Theorem 8. If a and b are in Z, so are $a +$ $a - b$.

Proof. We proceed in several steps.

A.

Step 1. We show that the theorem is true in the c where a is a positive integer and $b = 1$. That is, if a positive integer, we show that $a + 1$ and $a - 1$ are integers. That $a + 1$ is an integer (in fact, a positiv integer) has already been proved. We prove that $a - 1$ integer, by induction on a. It is true if $a = 1$, sinc $a - 1 = 0$ if $a = 1$. Supposing it true for a, we prov true for $a + 1$. That is, we show $(a+1) - 1$ is an inte But that is trivial, since $(a+1) - 1 = a$, which is an i by hypothesis (in fact, a positive integer).

Step 2. We show the theorem is true if a is any integer and $b = 1$.

We consider three cases. If a is a positive int this result follows from Step 1. If $a = 0$, the result immediate, since

$$0 + 1 = 1 \quad \text{and} \quad 0 - 1 = -1.$$

Finally, suppose $a = -c$, where c is a positive intege Then

$$a + 1 = -c + 1 = -(c-1),$$

$$a - 1 = -c - 1 = -(c+1).$$

5. Customers arrive at the Leftorium store according to a Poisson process with rate λ customers per hour. The true value of λ is unknown, so we treat it as a random variable (this is called a *Bayesian* approach). Suppose that our prior beliefs about λ can be expressed as $\lambda \sim \text{Expo}(3)$. Let X be the number of customers who arrive at the Leftorium between 1 pm and 3 pm tomorrow. Given that $X = 2$ is observed, find the conditional PDF of λ (this is known as the *posterior density* of λ).

Let us write Λ (capital λ) for the r.v. and λ for a specific possible value of Λ (in practice, often both would be written as λ, but to make sure the distinction between a r.v. and its values is clear we will maintain the notational distinction here). We need to find the conditional PDF of λ given the evidence, which we write as $f_{\Lambda|X}(\lambda|2)$. By Bayes' Rule, the posterior density of λ is

$$f_{\Lambda|X}(\lambda|x) = \frac{P(X = x|\Lambda = \lambda) f_\Lambda(\lambda)}{P(X = x)},$$

where $f_\Lambda(\lambda) = 3e^{-3\lambda}$ for $\lambda > 0$, and $P(X = x|\Lambda = \lambda)$ is found using the Pois(2λ) PMF. For $x = 2$, the numerator is

$$\frac{e^{-2\lambda}(2\lambda)^2}{2!} \cdot 3e^{-3\lambda} = 6\lambda^2 e^{-5\lambda}.$$

For the denominator when $x = 2$, note that it is a constant (not depending on λ), so it must be the constant that makes the conditional PDF integrate to 1. Equivalently, we can use the continuous version of the Law of Total Probability:

$$P(X = 2) = \int_0^\infty P(X = 2|\Lambda = \lambda) f_\Lambda(\lambda) d\lambda.$$

The conditional PDF is proportional to $\lambda^2 e^{-5\lambda}$, so we just need to find $\int_0^\infty \lambda^2 e^{-5\lambda} d\lambda$. This can be done using integration by parts, but a neater way is to recognize that

$$5 \int_0^\infty y^2 e^{-5y} dy = \frac{2!}{5^2} = \frac{2}{25},$$

as this is the integral we'd get using LOTUS to find $E(Y^2)$ for $Y \sim \text{Expo}(5)$. Thus,

$$f_{\Lambda|X}(\lambda|2) = \frac{125}{2} \lambda^2 e^{-5\lambda},$$

for $\lambda > 0$. (This is known as the Gamma(3,5) distribution.)

(d) If X and Y have the same distribution given Z, i.e., for all a and z, we have $P(X = a|Z = z) = P(Y = a|Z = z)$, then X and Y have the same distribution.

True: by the law of total probability, conditioning on Z gives

$$P(X = a) = \sum_z P(X = a|Z = z)P(Z = z).$$

Since X and Y have the same conditional distribution given Z, this becomes $\sum_z P(Y = a|Z = z)P(Z = z) = P(Y = a)$.

If X has MGF $M(t)$, what is the MGF of $-X$? What is the MGF of $a + bX$, where a and b are constants?

The MGF of $-X$ is $E(e^{-tX}) = M(-t)$. The MGF of $a + bX$ is

$$E(e^{t(a+bX)}) = E(e^{at+btX}) = e^{at}E(e^{btX}) = e^{at}M(bt).$$

A chicken lays n eggs. Each egg independently does or doesn't hatch, with probability p of hatching. For each egg that hatches, the chick does or doesn't survive (independently of the other eggs), with probability s of survival. Let $N \sim \text{Bin}(n, p)$ be the number of eggs which hatch, X be the number of chicks which survive, and Y be the number of chicks which hatch but don't survive (so $X + Y = N$). Find the marginal PMF of X, and the joint PMF of X and Y. Are they independent?

5

Marginally we have $X \sim \text{Bin}(n, ps)$, as shown on a previous homework problem using a story proof (the eggs can be thought of as independent Bernoulli trials with probability ps of success for each). Here X and Y are *not* independent,

7. Shakespeare wrote a total of 884647 words in his known works. Of course, many words are used more than once, and the number of distinct words in Shakespeare's known writings is 31534 (according to one computation). This puts a lower bound on the size of Shakespeare's vocabulary, but it is likely that Shakespeare knew words which he did not use in these known writings.

More specifically, suppose that a new poem of Shakespeare were uncovered, and consider the following (seemingly impossible) problem: give a good prediction of the number of words in the new poem that do not appear anywhere in Shakespeare's previously known works.

The statisticians Ronald Thisted and Bradley Efron studied this problem in a paper called "Did Shakespeare write a newly-discovered poem?", which performed statistical tests to try to determine whether Shakespeare was the author of a poem discovered by a Shakespearean scholar in 1985. A simplified version of their method is developed in the problem below. The method was originally invented by Alan Turing (the founder of computer science) and I.J. Good as part of the effort to break the German Enigma code during World War II.

Let N be the number of distinct words that Shakespeare knew, and assume these words are numbered from 1 to N. Suppose for simplicity that Shakespeare wrote only two plays, A and B. The plays are reasonably long and they are of the same length. Let X_j be the number of times that word j appears in play A, and Y_j be the number of times it appears in play B, for $1 \le j \le N$.

(a) Explain why it is reasonable to model X_j as being Poisson, and Y_j as being Poisson with the same parameter as X_j.

It is reasonable to model X_j and Y_j as Poisson, because this distribution is used to describe the number of "events" (such as emails received) happening at some average rate in a fixed interval or volume. The Poisson paradigm applies here: each

individual word in a play has some very small probability of being word j, and the words are weakly dependent. Here an event means using word j, the average rate is determined by how frequently Shakespeare uses that word overall. It is reasonable to assume that the average rate of occurrence of a particular word is the same for two plays by the same author, so we take λ to be the same for X_j and Y_j.

(b) Let the numbers of occurrences of the word "eyeball" (which was coined by Shakespare) in the two plays be independent Pois(λ) r.v.s. Show that the probability that "eyeball" is used in play B but not in play A is

$$e^{-\lambda}(\lambda - \lambda^2/2! + \lambda^3/3! - \lambda^4/4! + \ldots).$$

Let X be the number of times that "eyeball" is used in play A, and Y be the number of times that it is used in play B. Since X and Y are independent Pois(λ)

$$\begin{aligned} P(X=0, Y>0) &= P(X=0)\,(1-P(Y=0)) = e^{-\lambda}\left(1-e^{-\lambda}\right) \\ &= e^{-\lambda}\left(1-\left(1-\lambda+\frac{\lambda^2}{2!}-\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}-\ldots\right)\right) \end{aligned}$$

6. Fix $b > 1$.

(a) If m, n, p, q are integers, $n > 0, q > 0$, and $r = m/n = p/q$, prove that

$$(b^m)^{1/n} = (b^p)^{1/q}.$$

Hence it makes sense to define $b^r = (b^m)^{1/n}$.

(b) Prove that $b^{r+s} = b^r b^s$ if r and s are rational.

(c) If x is real, define $B(x)$ to be the set of all numbers b^t, where t is rational and $t \leq x$. Prove that

$$b^r = \sup B(r)$$

where r is rational. Hence, it makes sense to define

$$b^x = \sup B(x)$$

for every real x.

(d) Prove that $b^{x+y} = b^x b^y$ for every real x and y.

Solution:

(a) Frst observe that $(b^{n_1})^{n_2} = b^{n_1 n_2}$ for $n_1, n_2 \in \mathbb{Z}$ by direct expansion of either term. Let $y^n = b^m$ so that $y = (b^m)^{1/n}$. Then, $y^{nq} = b^{mq} = b^{np} \implies (y^q)^n = (b^p)^n$, because $m/n = p/q \implies mq = np$. By Theorem 1.12 on page 10, since n^{th} roots are unique, we have $y^q = b^p$. Then, $(b^m)^{1/n} = y = (b^p)^{1/q}$.

(b) Let $r, s \in \mathbb{Q}$ and write $r = m/n$ and $s = p/q$ for $m, n, p, q \in \mathbb{Z}$. Then, $r+s = m/n+p/q = (mq+np)/nq$. Note, for $n_1, n_2 \in \mathbb{Z}$ we have $b^{n_1} b^{n_2} = b^{n_1+n_2}$ by direct expansion. Then, by (a), $b^{r+s} = (b^{mq+np})^{1/nq} = (b^{mq} b^{np})^{1/nq} = (b^{mq})^{1/nq} (b^{np})^{1/nq}$, where the last equality follows from the Corollary to Theorem 1.12 on page 11 of the text. Again using the result of (a) we can simplify this last expression and obtain $(b^{mq})^{1/nq}(b^{np})^{1/nq} = (b^m)^{1/n}(b^p)^{1/q} = b^r b^s$, showing (b).

(c) Let $t, r \in \mathbb{Q}$ and write $t = p/q, r = m/n$, where $p, q, m, n \in \mathbb{Z}$. Then, $t < r \implies p/q < m/n \implies pn < mq$. Thus, since $b > 1$, we have that $b^{pn} < b^{mq} \implies b^p < (b^{mq})^{1/n} \implies (b^p)^{1/q} < ((b^{mq})^{1/n})^{1/q}$. Using the result of part (a), we can simplify the last expression to obtain $b^t = (b^p)^{1/q} < (b^m)^{1/n} = b^r$, showing that

$$b^r > \{b^t \mid t < r\} \implies b^r = \sup \{b^t \mid t \leq r\} = B(r)$$

and proving part (c).

(d) We first observe that $b^r b^t = b^{r+t}$ by part (b) and so

$$b^x b^y = \sup_{r \leq x, t \leq y} b^r b^t = \sup_{r \leq x, t \leq y} b^{r+t} \leq \sup_{r+t \leq x+y} b^{r+t} = b^{x+y}.$$

Here, $\sup_{r \leq x, t \leq y} b^{r+t} \leq \sup_{r+t \leq x+y} b^{r+t}$ because we are taking the sup over a more restricted set on the right since $r \leq x, t \leq y \implies r+t \leq x+y$ but $r+t \leq x+y$ doesn't imply $r \leq x$ and $t \leq y$. But, we in fact have equality because if $\sup_{r \leq x, t \leq y} b^{r+t} < \sup_{r+t \leq x+y} b^{r+t}$ then there exist $r', t' \in \mathbb{Q}$ such that $r' + t' \leq x + y$ and

$$\sup_{\rho \leq x, \tau \leq y} b^{\rho+\tau} < b^{r'+t'} \leq \sup_{r+t \leq x+y} b^{r+t}. \tag{6.1}$$

If $r' + t' = x + y$ we get a contradiction to Equation (6.1) by choosing $\rho = x, \tau = y$ (this follows from Exercise 1 above). Otherwise, $r' + t' < x + y$ and then we can choose $\rho < x, \tau < y$ such that $r' + t' < \rho + \tau < x + y$, because $\mathbb{Q}$ is dense in $\mathbb{R}$. But then, $b^{\rho+\tau} > b^{r'+t'}$, also contradicting Equation (6.1). Therefore we have equality and part (d) is shown. □

(c) For the case that X is continuous with PDF $f(x)$ which is positive everywhere, show that the value of c that minimizes $E|X-c|$ is the median of X (which is the value m with $P(X \leq m) = 1/2$.

Hint: this can be done either with or without calculus. For the calculus method, use LOTUS to write $E|X-c|$ as an integral, and then split the integral into 2 pieces to get rid of the absolute values. Then use the fundamental theorem of calculus (after writing, for example, $\int_{-\infty}^{c}(c-x)f(x)dx = c\int_{-\infty}^{c} f(x)dx - \int_{-\infty}^{c} xf(x)dx$).

Proof with calculus: We want to minimize

$$E|X-c| = \int_{-\infty}^{\infty} |x-c|f(x)dx = \int_{-\infty}^{c} (c-x)f(x)dx + \int_{c}^{\infty} (x-c)f(x)dx,$$

where we split the integral into 2 pieces to handle the absolute values. This becomes

$$E|X-c| = c\int_{-\infty}^{c} f(x)dx - \int_{-\infty}^{c} xf(x)dx + \int_{c}^{\infty} xf(x)dx - c\int_{c}^{\infty} f(x)dx.$$

Now differentiate both sides with respect to c, using the fundamental theorem of calculus:

$$\frac{d}{dc}(E|X-c|) = \int_{-\infty}^{c} f(x)dx + cf(c) - cf(c) - cf(c) - \int_{c}^{\infty} f(x)dx + cf(c),$$

which simplifies to $P(X \leq c) - (1 - P(X \leq c)) = 2P(X \leq c) - 1$. This is 0 when $P(X \leq c) = 1/2$. The second derivative is $f(c) + f(c) = 2f(c) > 0$, so we have found a minimum. Thus, $E|X-c|$ is minimized when c is the median.

Proof without calculus: We will show a more general result, not assuming that X is a continuous r.v. A number m is called a *median* of the distribution of X if $P(X \leq m) \geq 1/2, P(X \geq m) \geq 1/2$. (So a median may not be unique; it will be unique if the CDF is continuous and strictly increasing.) Let m be a median, and let $a \neq m$. We need to show $E|X-m| \leq E|X-a|$, which is equivalent to $E(|X-a| - |X-m|) \geq 0$. Assume $m < a$ (the case $m > a$ can be handled by the same method).

Note that $|X-a| - |X-m| = a - X - (m - X) = a - m$ if $X \leq m$, and $|X-a| - |X-m| \geq X - a - (X - m) = m - a$ if $X > m$. Splitting the definition of expected value into 2 parts based on whether $X > m$ occurs, we have

$$E(|X-a| - |X-m|) \geq (a-m)P(X \leq m) + (m-a)P(X > m),$$

which simplifies to $(a-m)(P(X \leq m) - P(X > m))$. By definition of median,

$$P(X \leq m) - P(X > m) = P(X \leq m) - (1 - P(X \leq m)) = 2P(X \leq m) - 1 \geq 0,$$

which shows that $E(|X-m|) \leq E(|X-a|)$.

There are 100 passengers lined up to board an airplane with 100 seats (with each seat assigned to one of the passengers). The first passenger in line crazily decides to sit in a randomly chosen seat (with all seats equally likely). Each subsequent passenger takes his or her assigned seat if available, and otherwise sits in a random available seat. What is the probability that the last passenger in line gets to sit in his or her assigned seat? (This is another common interview problem, and a beautiful example of the power of symmetry.)

4

Hint: call the seat assigned to the jth passenger in line "Seat j" (regardless of whether the airline calls it seat 23A or whatever). What are the possibilities for which seats are available to the last passenger in line, and what is the probability of each of these possibilities?

The seat for the last passenger is either Seat 1 or Seat 100; for example, Seat 42 can't be available to the last passenger since the 42nd passenger in line would have sat there if possible. Seat 1 and Seat 100 are equally likely to be available to the last passenger, since the previous 99 passengers view these two seats symmetrically. So the probability that the last passenger gets Seat 100 is 1/2.

(b) Three legs are positioned uniformly and independently on the perimeter of a round table. What is the probability that the table will stand?

Think of the legs as points on a circle, chosen randomly one at a time, and choose units so that the circumference of the circle is 1. Let A, B, C be the arc lengths from one point to the next (clockwise, starting with the first point chosen). Then

$$\begin{aligned} P(\text{table falls}) &= P(\text{the 3 legs are all contained in some semicircle}) \\ &= P(\text{at least one of } A, B, C \text{ is greater than } 1/2) = 3/4, \end{aligned}$$

by Part (a). So the probability that the table will stand is $1/4$.

Alternatively, let C_j be the clockwise semicircle starting from the jth of the 3 points. Let A_j be the event that C_j contains all 3 points. Then $P(A_j) = 1/4$ and with probability 1, at most one A_j occurs. So $P(A_1 \cup A_2 \cup A_3) = 3/4$, which again shows that the probability that the table will stand is $1/4$.

---

4. Let F be a CDF which is continuous and strictly increasing. The inverse function, F^{-1}, is known as the *quantile function*, and has many applications in statistics and econometrics. Find the area under the curve of the quantile function from 0 to 1, in terms of the mean μ of the distribution F. Hint: Universality.

We want to find $\int_0^1 F^{-1}(u)du$. Let $U \sim \text{Unif}(0, 1)$ and $X = F^{-1}(U)$. By Universality of the Uniform, $X \sim F$. By LOTUS,

$$\int_0^1 F^{-1}(u)du = E(F^{-1}(U)) = E(X) = \mu.$$

Equivalently, make the substitution $u = F(x)$, so $du = f(x)dx$, where f is the PDF of the distribution with CDF F. Then the integral becomes

$$\int_{-\infty}^{\infty} F^{-1}(F(x))f(x)dx = \int_{-\infty}^{\infty} xf(x)dx = \mu.$$

n
