Ravi Dayabhai

offers are i.i.d $X_1, X_2, \ldots$ such that
$X_i \sim \text{Expo}(.0001) \Rightarrow P(X_i \leq x) = F(x)$,
$1 - F(x) = P(X_i \geq x) = 1 - (1 - e^{-.0001x}) \Rightarrow P(X_i \geq 15000) = e^{-(.0001)15000}$
This is the probability of getting a satisfactory bid.

# Stat 110 Strategic Practice 6, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

## 1 Exponential Distribution and Memorylessness

1. Fred (the protagonist of HW 6 #1) wants to sell his car, after moving back to Blissville (where he is happy with the bus system). He decides to sell it to the first person to offer at least \$15,000 for it. Assume that the offers are independent Exponential random variables with mean \$10,000.

   (a) Find the expected number of offers Fred will have. (SEE ATTACH.)

   (b) Find the expected amount of money that Fred gets for the car. (SEE ATTACH.)

2. Find $E(X^3)$ for $X \sim \text{Expo}(\lambda)$, using LOTUS and the fact that $E(X) = 1/\lambda$ and $\text{Var}(X) = 1/\lambda^2$, and integration by parts at most once (see also Problem 1 in the MGF section).

3. Let $X_1, \ldots, X_n$ be independent, with $X_j \sim \text{Expo}(\lambda_j)$. (They are i.i.d. if all the $\lambda_j$'s are equal, but we are not assuming that.) Let $M = \min(X_1, \ldots, X_n)$. Show that $M \sim \text{Expo}(\lambda_1 + \cdots + \lambda_n)$, and interpret this intuitively.

4. A post office has 2 clerks. Alice enters the post office while 2 other customers, Bob and Claire, are being served by the 2 clerks. She is next in line. Assume that the time a clerk spends serving a customer has the Exponential($\lambda$) distribution.

   (a) What is the probability that Alice is the last of the 3 customers to be done being served Hint: no integrals are needed.

   (b) What is the expected total time that Alice needs to spend at the post office?

## 2 Moment Generating Functions (MGFs)

1. Find $E(X^3)$ for $X \sim \text{Expo}(\lambda)$ using the MGF of $X$ (see also Problem 2 in the Exponential Distribution section).

2. 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?



*SEE ATTACH for remainder of SP 6.

[SP 6, Exponential Distribution & Memorylessness]

① A $p = P(X_i \geq 15000) = e^{-(0.0001)15000} = e^{-1.5}$

Let $Y \sim Geom(p)$, $Y$ counts the number of failed bids before the first successful bid.

$1 + E(Y)$ = expected # of offers (including the successful one)

$= 1 + \frac{q}{p} = \cancel{1} + \frac{1 - e^{-1.5}}{e^{-1.5}} = \frac{1}{e^{-1.5}} = \boxed{e^{1.5}}$ offers, on average

B The expected amount of money is $E(X \mid X \geq 15000)$, so by MEMORYLESSNESS, $15000 + 10\,000 = \boxed{25\,000}$

($X$ here is the offer that hit the bid.)

② $X \sim Expo(\lambda)$; use LOTUS to find $E(X^3)$ — 3rd moment of $X$.

$$E(X^3) = \int_0^\infty x^3 \lambda e^{-\lambda x}\, dx = uv\Big|_0^\infty - \int_0^\infty v\, du = \cancel{-x^3 e^{-\lambda x}}\Big|_0^\infty + \int_0^\infty 3x^2 e^{-\lambda x}\, dx$$

$u = x^3$ $\quad dv = \lambda e^{-\lambda x}\, dx$

$du = 3x^2\, dx$ $\quad v = \int_0^\infty \lambda e^{-\lambda x}\, dx = -e^{-\lambda x}\Big|_0^\infty = 1$

integration by parts

$$= \frac{3}{\lambda}\int_0^\infty x^2 \lambda e^{-\lambda x}\, dx = \frac{3}{\lambda} E(X^2) = \frac{3}{\lambda}\left(Var(X) + E(X)^2\right) = \frac{3}{\lambda}\cdot\frac{2}{\lambda^2} = \boxed{\frac{6}{\lambda^3}}$$

③ $M = \min(X_1, \ldots, X_n)$ where $X_j \sim Expo(\lambda_j)$ and all $X_j$ are independent. Show that $M \sim Expo(\lambda_1 + \lambda_2 + \ldots + \lambda_n)$.

$F_M(m) = P(\min(X_1, \ldots, X_n) < m)$. This implies at least 1 $X_j < m$, which is equivalent to the complement of all $X_j > m$.

$$F_M(m) = 1 - \prod_j 1 - F_{X_j}(m) = 1 - e^{-(\lambda_1 + \lambda_2 + \ldots + \lambda_n)m} \Rightarrow M \sim Expo\left(\sum_j \lambda_j\right)$$

(by independence)

④ A $A, B, C \sim Expo(\lambda)$. Let $|b - c| = t$; we want to know $P(A \geq t) = P(A \geq s + t \mid A \geq s)$ where $s < \min(b, c)$, which is the memoryless property of the Expo distribution.

Directly, $P(A \geq t) = 1 - (1 - e^{-\lambda t}) = \boxed{e^{-\lambda t}}$

Use symmetry of 2 i.i.d cont. r.v.s since when Alice starts to get served, remaining person's time is also $Expo(\lambda)$.

$\Rightarrow \boxed{\frac{1}{2}}$

(SP 6, Exponential Distribution & Memorylessness, CONT.)

④ B. Using same conventions from A, Alice's expected total time is $E(A \mid A \geq s) = s + \frac{1}{\lambda}$.

Yes, but we don't know time spent in line!
Use fact from #3 that $\min(B, C) \sim \text{Expo}(2\lambda)$, so instead of $s$, $\frac{1}{2\lambda}$. Then expected total time is
$\frac{1}{2\lambda} + \frac{1}{\lambda} = \boxed{\frac{3}{2\lambda}}$.

[SP 6, Moment Generating Functions]

① $X \sim \text{Expo}(\lambda)$. $E(X^3) = ?$ Letting $M(t) = E(e^{tX})$, the moment-generating function for $X$:

$$M(t) = \sum_{n=0}^{\infty} M^{(n)}(0) \frac{t^n}{n!} \Rightarrow E(X^3) = M^{(3)}(0).$$

$$= E(e^{tX}) = \lambda \int_0^{\infty} e^{tu} \cdot e^{-\lambda u}\, du = \lambda \int_0^{\infty} e^{(t-\lambda)u}\, du = \frac{\lambda}{t-\lambda} e^{(t-\lambda)u} \Big|_0^{\infty}$$

When $t < \lambda$, this evaluates to $-\frac{\lambda}{t-\lambda} = \frac{\lambda}{\lambda - t}$ and $M(0) = 1$
(i.e., think $E(e^{tX})$ is finite for $t \in (-a, a)$ where $a \leq \lambda$).

$M^{(1)}(t) = \lambda(\lambda - t)^{-2}$, $M^{(2)}(t) = +2\lambda(\lambda - t)^{-3}$, $\boxed{M^{(3)}(t) = 6\lambda(\lambda - t)^{-4} \quad M^{(3)}(0) = \frac{6}{\lambda^3}}$

② Taylor Series of $e^{-x}$ @ $x = 0$ $= 1 - x + \frac{x^2}{2} - \frac{x^3}{6} + \frac{x^4}{24} - \frac{x^5}{120} + \ldots$

$$\boxed{M(-t) = \sum_{n=0}^{\infty} (-1)^n \frac{t^n}{n!} M^{(n)}(0) = \sum_{n=0}^{\infty} (-1)^n E(X^n) \frac{t^n}{n!} = E(e^{-tX})}$$

$E(e^{t(a+bX)}) = E(e^{ta} e^{tbX}) = e^{ta} E(e^{tbX}) = \boxed{e^{ta} M(bt)}$ is MGF for $a + bX$; $a, b$ are constants.

③ $X = U_1 + U_2 + \ldots + U_{60}$, where $U_i \sim \text{Unif}(0,1)$. MGF of $X$: $E(e^{tX})$,
so $E(e^{t(U_1 + \ldots + U_{60})}) = E(e^{tU_1} e^{tU_2} e^{tU_3} \cdots e^{tU_{60}})$. Because all $U_i$ are i.i.d,
this is equivalent to $E(e^{tU_1}) E(e^{tU_2}) \ldots E(e^{tU_{60}}) = E(e^{tU})^{60}$.

$$E(e^{tU})^{60} = M(t)^{60} = \left[\frac{1}{1-0} \int_0^1 e^{tu}\, du\right]^{60} = \boxed{\left(\frac{e^t - 1}{t}\right)^{60}}$$

3. Let $U_1, U_2, \ldots, U_{60}$ be i.i.d. Uniform(0,1) and $X = U_1 + U_2 + \cdots + U_{60}$. Find the MGF of $X$.

4. Let $X \sim \text{Pois}(\lambda)$, and let $M(t)$ be the MGF of $X$. The *cumulant generating function* is defined to be $g(t) = \ln M(t)$. Expanding $g(t)$ as a Taylor series

$$g(t) = \sum_{j=1}^{\infty} \frac{c_j}{j!} t^j$$

(the sum starts at $j = 1$ because $g(0) = 0$), the coefficient $c_j$ is called the $j$th *cumulant* of $X$. Find the $j$th cumulant of $X$, for all $j \geq 1$.

---

(HW 6) ALMOST; "bad" math! Correct approach.

② $X_1, X_2, X_3 \sim \text{Expo}\left(\frac{1}{6}\right)$. $E(\max(X_1, X_2, X_3)) = ?$ By SYMMETRY, $P(X_1 < X_2 < X_3) = P(X_2 < X_1 < X_3) = P(X_3 < X_2 < X_1) = \ldots = \frac{1}{6}$. We need the CDF of $\max(X_1, X_2, X_3) = M$, $F_M(m) = P(\max(X_1,X_2,X_3) < m) \Rightarrow$ $P(\text{all } X_i < m \text{ for } i \in \{1,2,3\})$. By INDEPENDENCE, $F_M(m) = \prod_{i=3}^{3} F_{X_i}(m) = (1-e^{-\frac{m}{6}})^3$. The PDF of $M = f(m) = F_M'(m) = \frac{d}{dm}(1-e^{-\frac{m}{6}})^3 =$ $3(1-e^{-\frac{m}{6}})^2 \cdot (-e^{-\frac{m}{6}}) \cdot (-\frac{1}{6}) = \underbrace{\frac{1}{2}e^{-\frac{m}{6}}(1-e^{-\frac{m}{6}})^2}$ $=$ , Then $E(M) = \int_0^\infty m \cdot \boxed{\phantom{x}} \, dm$, which is $\boxed{11 \text{ hours.}}$ $\leftarrow$ ~~This is correct if $M = \max(X_1, X_2, X_3) \ldots$ but the problem is~~ $T = T_1 + T_2 + T_3$; $T_1 = \min(X_1, X_2, X_3)$, $T_2$ is is additional time for a second student to finish, by MEMORYLESSNESS, $T_2 \sim \text{Expo}\left(\frac{2}{6}\right)$; similarly for $T_3 \sim \text{Expo}\left(\frac{1}{6}\right)$. $\boxed{E(T) = 11 \text{ hours}}$.

My answer is the same! Memorylessness is key to establishing that $M = T$!

## Stat 110 Homework 6, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. Fred lives in Blissville, where buses always arrive exactly on time, with the time between successive buses fixed at 10 minutes. Having lost his watch, he arrives at the bus stop at a random time (assume that buses run 24 hours a day, and that the time that Fred arrives is uniformly random on a particular day). (SEE ATTACH)

(a) What is the distribution of how long Fred has to wait for the next bus? What is the average time that Fred has to wait?

(b) Given that the bus has not yet arrived after 6 minutes, what is the probability that Fred will have to wait at least 3 more minutes?

(c) Fred moves to Blotchville, a city with inferior urban planning and where buses are much more erratic. Now, when any bus arrives, the time until the next bus arrives is an Exponential random variable with mean 10 minutes. Fred arrives at the bus stop at a random time, not knowing how long ago the previous bus came. What is the distribution of Fred's waiting time for the next bus? What is the average time that Fred has to wait? (Hint: don't forget the memoryless property.)

(d) When Fred complains to a friend how much worse transportation is in Blotchville, the friend says: "Stop whining so much! You arrive at a uniform instant between the previous bus arrival and the next bus arrival. The average length of that interval between buses is 10 minutes, but since you are equally likely to arrive at any time in that interval, your average waiting time is only 5 minutes."

Fred disagrees, both from experience and from solving Part (c) while waiting for the bus. Explain what (if anything) is wrong with the friend's reasoning.

2. Three Stat 110 students are working independently on this pset. All 3 start at 1 pm on a certain day, and each takes an Exponential time with mean 6 hours to complete this pset. What is the earliest time when all 3 students will have completed this pset, on average? (That is, all of the 3 students need to be done with this pset.) (SEE PREVIOUS PAGE)

3. Consider an experiment where we observe the value of a random variable $X$, and estimate the value of an unknown constant $\theta$ using some random variable $T = g(X)$ that is a function of $X$. The r.v. $T$ is called an *estimator*. Think of $X$ as the data observed in the experiment, and $\theta$ as an unknown parameter related to the distribution of $X$.

(SP 6, MGFs, CONT.)

If $X \sim Pois(\lambda)$, then the $MGF_X(t) = E(e^{tX}) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x e^{tx}}{x!}$.

This simplifies to $e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$.

The cumulant generating function $g(t) = \ln MGF_X(t)$, which is $\ln e^{\lambda(e^t-1)} = \lambda(e^t - 1)$.

$e^t = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \ldots$, so $g_X(t) = \lambda \sum_{j=1}^{\infty} \frac{t^j}{j!} \Rightarrow \boxed{\lambda}$ is the $j^{th}$ cumulant of $X$ for all $j \geq 1$.

[HW 6]

① $\boxed{a}$ $X \equiv$ waiting time for bus; in Blissville $\boxed{X \sim Unif(0, 10)}$

Fred's average waiting time is $E(X) = \int_0^{10} \frac{x}{10}\, dx$ (by LOTUS)

$= \frac{1}{20} x^2 \Big|_0^{10} = \boxed{5 \text{ mins.}}$

$\boxed{b}$. $P(X \geq 9 \mid X \geq 6) = \frac{P(X \geq 9)}{P(X \geq 6)} = \frac{1 - P(X \leq 9)}{1 - P(X < 6)} = \frac{0.1}{0.4}$

$= \boxed{0.25}$

$\boxed{c}$ $Y \equiv$ waiting time for bus; in Blotchville $Y \sim Expo(\frac{1}{10})$.

Irrespective of when Fred arrives at the bus stop, his waiting time is distributed $Expo(\frac{1}{10})$ by the MEMORYLESS PROPERTY. Similarly his average wait time is still 10 mins.

ALMOST $\boxed{d}$. The average inter-arrival time might be 10 mins., but Fred arrives randomly between particular intervals which is equally likely to mean Fred's wait is ~~$0 \leq$~~ $y \leq 10$ mins. or $y \geq 0$, the latter of which increases his wait time in Blotchville relative to Blissville (i.e., long-tail risk). Fred's friend is confusing expectation with the distribution.

\+ Fred is not equally likely to arrive in a short interval vs. long interval between buses! This is length biasing

For example, consider the experiment of flipping a coin $n$ times, where the coin has an unknown probability $\theta$ of Heads. After the experiment is performed, we have observed the value of $X \sim \text{Bin}(n, \theta)$. The most natural estimator for $\theta$ is then $X/n$.

(a) The *bias* of an estimator $T$ for $\theta$ is defined as $b(T) = E(T) - \theta$. The *mean squared error* is the average squared error when using $T(X)$ to estimate $\theta$:

$$\text{MSE}(T) = E(T - \theta)^2.$$

Show that

$$\text{MSE}(T) = \text{Var}(T) + (b(T))^2.$$

This implies that for fixed MSE, lower bias can only be attained at the cost of higher variance and vice versa; this is a form of the *bias-variance tradeoff*, a phenomenon which arises throughout statistics.

(b) Show without using calculus that the constant $c$ that minimizes $E(X - c)^2$ is the expected value of $X$. (This means that in choosing a single number to summarize $X$, the mean is the best choice if the goal is to minimize the average squared error.)

(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 $P(X \le m) = 1/2$. $= P(X > m) = F(m) = \int_{-\infty}^{m} f(x)\,dx$

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, $\int_{-\infty}^{c}(c - x)f(x)dx = c\int_{-\infty}^{c} f(x)dx - \int_{-\infty}^{c} xf(x)dx$).

4. (a) Suppose that we have a list of the populations of every country in the world.

30%

*Guess*, without looking at data yet, what percentage of the populations have the digit 1 as their first digit (e.g., a country with a population of 1,234,567 has first digit 1 and a country with population 89,012,345 does not).

**Note**: (a) is a rare problem where the only way to lose points is to find out the right answer rather than guessing! LOL !

(b) After having done (a), look through a list of populations such as

http://en.wikipedia.org/wiki/List_of_countries_by_population

and count how many start with a 1. What percentage of countries is this?

$$\boxed{\frac{73}{241} \doteq 30.3\%}$$

(HW 6, CONT.)

5A. Show that $b(T) = E(T) - \theta$, bias of estimator $T$ for $\theta$, $MSE(T) = E(T-\theta)^2 \implies MSE(T) = Var(T) + b(T)^2$ (this is a form of the bias-variance trade-off!)

$$MSE(T) = E(T^2 - 2T\theta + \theta^2) = E(T^2) - 2\theta E(T) + \theta^2$$

(complete the square)

$$= E(T-\theta)^2 + E(T)^2 = E(T^2) + \theta^2 - 2\theta E(T) + E(T)^2$$

$$= MSE(T) = E(T-\theta)^2 = \underbrace{E(T^2) - E(T)^2}_{Var(T)} + \underbrace{(E(T) - \theta)^2}_{b(T)^2}$$

B. Show that $c$ which minimizes $E(X-c)^2$ is $E(X)$ without calculus: $E(X^2 - 2Xc + c^2) = E(X^2) - 2cE(X) + c^2$. Like above, we can complete the square and see that $E(X-c)^2 = E(X^2) - E(X)^2 + (E(X) - c)^2$. The variance of $X$ is invariant with regards to $c$, so $(E(X)-c)^2$ is minimized when $c = E(X) \Rightarrow E(X-c)^2$ is minimized when $c = E(X)$.

C. For continuous r.v. $X$ with PDF $f(x)$ positive everywhere, show that $c$ which minimizes $E|X-c|$ is the median of $X$ (i.e., $m$ such that $P(X \le m) = \frac{1}{2}$).

$$\underset{c \in (-\infty,\infty)}{\arg\min}\, E|X-c| = \underset{c \in (-\infty,\infty)}{\arg\min} \left( \int_{-\infty}^{\infty} |x-c| f(x)\, dx \right) = \underset{c \in (-\infty,\infty)}{\arg\min} \left( \int_{-\infty}^{c} (c-x) f(x)\, dx + \int_{c}^{\infty} (x-c) f(x)\, dx \right)$$

$$= \underset{c \in (-\infty,\infty)}{\arg\min} \underbrace{\left( c\left[ \int_{-\infty}^{c} f(x)\, dx - \int_{c}^{\infty} f(x)\, dx \right] + \left[ \int_{c}^{\infty} x f(x)\, dx - \int_{-\infty}^{c} x f(x)\, dx \right] \right)}$$

$$\cancel{= \underset{c \in (-\infty,\infty)}{\arg\min}\; c(2F(c) - 1) + E(X) - 2\int_{-\infty}^{c} x f(x)\, dx} \qquad \frac{d}{dx}$$

$$\cancel{= \underset{c \in (-\infty,\infty)}{\arg\min} \left( c(2F(c) - 1) \right)}$$

$\Rightarrow c = m$ minimizes $E|X-c|$ because $P(X \le c) - (1 - P(X \le c)) = 2P(X \le c) - 1 = 0 \implies c = m$. The second derivative of the above expression is $f(c) + f(c) = 2f(c) > 0$, so we know $m$ is a minimum.

$P(D=j) = \log_{10}\left(\frac{j+1}{j}\right)$ is a valid PMF because it is non-negative (log function,

and $\sum_{j=1}^{9} \log_{10}\left(\frac{j+1}{j}\right) = \log_{10}\left(\frac{\cancel{1}+1}{1} \cdot \frac{\cancel{2}+1}{\cancel{2}} \cdot \frac{3+1}{\cancel{3}} \cdots \frac{9+1}{\cancel{9}}\right) = \log_{10}(10) = 1$.

(c) *Benford's Law* states that in a very large variety of real-life data sets, the first digit approximately follows a particular distribution with about a 30% chance of a 1, an 18% chance of a 2, and in general

$\log_{10}\left(1+\frac{1}{j}\right) \approx \frac{1}{j}$ for big $j$

$$P(D=j) = \log_{10}\left(\frac{j+1}{j}\right), \text{ for } j \in \{1,2,3,\ldots,9\},$$

where $D$ is the first digit of a randomly chosen element. Check that this is a PMF (using properties of logs, not with a calculator). (SEE ABOVE)

(d) Suppose that we write the random value in some problem (e.g., the population of a random country) in scientific notation as $X \times 10^N$, where $N$ is a nonnegative integer and $1 \leq X < 10$. Assume that $X$ is a continuous r.v. with PDF

$$f(x) = c/x, \text{ for } 1 \leq x \leq 10$$

(SEE BACK)

(and 0 otherwise), with $c$ a constant. What is the value of $c$ (be careful with the bases of logs)? Intuitively, we might hope that the distribution of $X$ does not depend on the choice of units in which $X$ is measured. To see whether this holds, let $Y = aX$ with $a > 0$. What is the PDF of $Y$ (specifying where it is nonzero)?

(e) Show that if we have a random number $X \times 10^N$ (written in scientific notation) and $X$ has the PDF $f(x)$ from (d), then the first digit (which is also the first digit of $X$) has the PMF given in (c). (SEE BACK)

Hint: what does $D = j$ correspond to in terms of the values of $X$?

5. Customers arrive at the Leftorium store according to a Poisson process with rate $\lambda$ customers per hour. The true value of $\lambda$ is unknown, so we treat it as a random variable (this is called a *Bayesian* approach). Suppose that our prior beliefs about $\lambda$ 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 $\lambda$ (this is known as the *posterior density* of $\lambda$).

6. Let $X_n \sim \text{Bin}(n, p_n)$ for all $n \geq 1$, where $np_n$ is a constant $\lambda > 0$ for all $n$ (so $p_n = \lambda/n$). Let $X \sim \text{Pois}(\lambda)$. Show that the MGF of $X_n$ converges to the MGF of $X$ (this gives another way to see that the $\text{Bin}(n,p)$ distribution can be well-approximated by the $\text{Pois}(\lambda)$ when $n$ is large, $p$ is small, and $\lambda = np$ is moderate). (SEE BACK)

7. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = e^X$. Then $Y$ has a *Log-Normal distribution* (which means "log is Normal"; note that "log of a Normal" doesn't make sense since Normals can be negative).

Find the mean and variance of $Y$ using the MGF of $X$, *without doing any integrals.* Then for $\mu = 0, \sigma = 1$, find the $n$th moment $E(Y^n)$ (in terms of $n$).

4. D In order to be a valid PDF, $c\int_1^{10} \frac{1}{x}\, dx = 1$. We know $c$ must be positive since the integral is positive over $(1, 10)$.

$c\int_1^{10} \frac{1}{x} = c\left[\ln(10) - \ln(1)\right] = c \ln 10 = 1 \Rightarrow \boxed{c = \frac{1}{\ln 10}}$

If we let $Y = aX$ for $a > 0$, then we can find the PDF of $Y$. $\int_a^{10a} f(y)\, dy = 1$ must be confirmed. If we let $f(y) = \frac{c}{y}$, $c\int_a^{10a} \frac{1}{y}\, dy = c \cdot \ln(10) = 1$, so the PDF of $Y$ is the same as the PDF of $X$, but over $[a, 10a]$ instead of $[1, 10]$.

E Let $X$ have PDF $f(x) = \frac{c}{x}$ where $c = \frac{1}{\ln 10} = \log_{10} e$. If we ~~in~~ this continuous r.v. we can show that it corresponds to PMF from C: bins are $[1,2), [2,3), \ldots, [9,10)$. We can show that $\int_j^{j+1} \frac{1}{x \ln 10}\, dx = \log_{10}\left(\frac{j+1}{j}\right)$ for $j \in \{1, 2, \ldots, 9\}$. Evaluating LHS gives $\frac{1}{\ln 10}\int_j^{j+1} \frac{1}{x}\, dx = \frac{\ln\left(\frac{j+1}{j}\right)}{\ln 10}$. By CHANGE OF BASE, we get

$$\frac{\log_{10}\left(\frac{j+1}{j}\right)}{\cancel{\log_{10}(e)}} \cdot \frac{\cancel{\log_{10}(e)}}{\log_{10}(10)} = \log_{10}\left(\frac{j+1}{j}\right).$$

6 MGF of $X \sim \text{Pois}(\lambda)$: $E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!}$

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

MGF of $X_n \sim \text{Bin}(n, p_n)$, $\lambda = np_n$ is constant, $\lambda > 0$:

$$E(e^{tX_n}) = (p_n e^t + q)^n = \left(\frac{\lambda e^t}{n} + \frac{n - \lambda}{n}\right)^n = \left(\frac{\lambda(e^t - 1)}{n} + 1\right)^n$$

Letting $n \to \infty$, $\left(1 + \frac{\lambda(e^t - 1)}{n}\right)^n = \boxed{e^{(e^t - 1)\lambda}}$.

[HW #6 (CONT.)]

⑤ PRIORS : $\lambda \sim \text{Expo}(3)$. We observe $X = 2$ from 1-3 pm

Find conditional PDF of $\lambda$ (POSTERIOR). $\lambda = \frac{\text{customers}}{\text{hour}}$

PRIOR PDF : $3e^{-3\lambda} = f(\lambda)$ ; PRIOR CDF : $1 - e^{-3\lambda} = F(\lambda)$

condition on having observed that $X = 2$ from 1-3 PM

(i.e. $\lambda = 1$?) → This maximizes likelihood of $X=2$

$X$ counts # of customers arrived between 1-3 pm

(2 hours)

→ $X \sim \text{Pois}(2\lambda)$ from Poisson process.

$$P(X=2) = \frac{e^{-2\lambda}(2\lambda)^{2}}{2} = 2e^{-2\lambda}\lambda^2 = 2(e^{-\lambda})^2\lambda^2 = 2(\lambda e^{-\lambda})^2$$

(PRIOR) PDF of $\lambda \sim \text{Expo}(3)$ vs. (POSTERIOR) PDF of $\lambda$ | ($X \sim \text{Pois}(2\lambda)$) $X = 2$

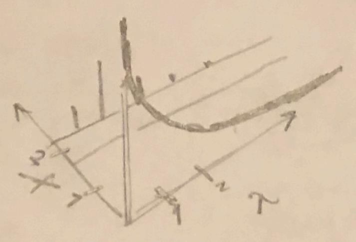
$$\frac{p(x \cap \theta)}{p(x)} = p(\theta | x) = \frac{p(x|\theta)\, p(\theta)}{p(x)}$$

?

$$\frac{p(x \cap \theta)}{p(\theta)}$$

$3e^{-3\lambda}$

What is joint distribution function for $X \sim \text{Pois}(2\lambda)$ & $\lambda \sim \text{Expo}(3)$ ?




CORRECT REASONING! But did not recognize that numerator is $6e^{-5\lambda}\lambda^2$

$$f_{\Lambda|X}(\lambda|2) = \frac{P(X=2|\Lambda=\lambda)\; f_\Lambda(\lambda)}{P(X=2)}$$ by Bayes Rule

$$= \frac{2e^{-2\lambda}\lambda^2 \cdot 3e^{-3\lambda}}{P(X=2)} = \frac{6e^{-5\lambda}\lambda^2}{P(X=2)}$$ by Poisson Process

normalizing constant

$$= \frac{6e^{-5\lambda}\lambda^2}{\int_0^\infty P(X=2|\Lambda=\lambda)\, f_\Lambda(\lambda)\, d\lambda}$$ by LOTP (continuous version)

$$1 = \int_0^\infty \frac{6e^{-5\lambda}\lambda^2}{P(X=2)}\, d\lambda \iff P(X=2) = \int_0^\infty 6e^{-5\lambda}\lambda^2\, d\lambda$$

$$= \frac{\cancel{6}e^{-5\lambda}\lambda^2}{\cancel{6}\int_0^\infty e^{-5\lambda}\lambda^2\, d\lambda} = \boxed{\frac{125}{2}e^{-5\lambda}\lambda^2}$$

by LOTUS

$$E(Y^2) = 5\int_0^\infty y^2 e^{-5y}\, dy = \frac{2}{5^2} = \frac{2}{25},$$ so divide by $5 \Rightarrow \frac{2}{125}$

for $Y \sim \text{Expo}(5)$!

CHECK via INTEGRATION BY PARTS:

$$\int_0^\infty \underbrace{\lambda^2}_{u}\underbrace{e^{-5\lambda}\, d\lambda}_{dv} = uv\Big|_0^\infty - \int_0^\infty v\, du = \cancel{-\frac{\lambda^2 e^{-5\lambda}}{5}\Big|_0^\infty} + \frac{2}{5}\underbrace{\int_0^\infty \lambda e^{-5\lambda}\, d\lambda}$$

again

$$\cancel{-\frac{\lambda e^{-5\lambda}}{5}\Big|_0^\infty} + \frac{1}{5}\int_0^\infty e^{-5\lambda}\, d\lambda \checkmark$$

$u = \lambda^2 \iff du = 2\lambda\, d\lambda$

$dv = e^{-5\lambda}\, d\lambda \iff \int dv = v = \int e^{-5\lambda}\, d\lambda$

$= -\frac{1}{5}e^{-5\lambda}$

⑦ ~~MGF~~ of $X$ : $E(e^{tX})$ but if we let $Y = a + bX$ (scale-location transformation), the $E(e^{t(a+bX)}) = E(e^{ta} e^{tbX}) = e^{ta} E(e^{tbX}) = M(bX)$. *

For standard normal r.v. $Z$, $M_Z(t) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz$.

"Completing the square" yields $M_Z(t) = \int_{-\infty}^{\infty} e^{t^2/2} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} \, dz$.

Simplifying, $M_Z(t) = e^{\frac{t^2}{2}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} \, dz}_{=1 \text{ because it is a valid PDF}}$, so $M_Z(t) = e^{t^2/2}$ **

let $X \sim N(\mu, \sigma^2)$, $M_X(t) = e^{t\mu} e^{(\sigma t)^2/2} = e^{t(\mu - \frac{t\sigma^2}{2})}$ by * & ** above. So, if

$Y = e^X \Leftrightarrow Y^n = e^{nX}$

$\Updownarrow$

$$E(Y^n) = E(e^{nX}) = \boxed{M_Y(n) = e^{\mu n + \frac{1}{2}\sigma^2 n^2}} \Rightarrow \boxed{\text{for } \mu = 0, \sigma = 1: \; E(Y^n) = E(Z) = e^{n^2/2}}$$

$E(Y) = \boxed{e^{\mu + \frac{1}{2}\sigma^2}}$ ✓

$Var(Y) = E(Y^2) - E(Y)^2 = e^{2\mu + 2\sigma^2}$ ✓ $- \left[e^{\mu + \frac{1}{2}\sigma^2}\right]^2 = \boxed{e^{2(\mu+\sigma^2)} - e^{2\mu + \sigma^2}}$ ✓