1 Means, medians, modes, and moments

1. The CDF is $F(x) = \frac{x-a}{b-a}$ for $x \in [a, b]$ and $F(x) = 1/2$ when $x = (a+b)/2$. The pdf is $\frac{1}{b-a}$ which is constant, so every $x \in [a, b]$ is a mode of U.

2. $F(x) = 1 - e^{-\lambda x} = 1/2 \implies \ln(2)/\lambda = x$ is the median. The PDF is $f(x) = \lambda e^{-\lambda x}$ which is maximized at $x = 0$, so the mode is $x = 0$.

3. The CDF is $P(X \leq x) = \int_1^x a/x^{a+1}dx = -1/x^a|_1^x = 1-x^{-a}$. $F(x) = 1/2 \implies x = 2^{1/a}$ is the median. The maximum is at $x = 1$ so the mode is $x = 1$.

4. (a) The mean is $np = 5/3$. Any number between 1 and 2 is the median and 2 is the unique mode.

   (b) The mean is 2. 2 is the unique median and mode.

5. If n is odd, then the median is $(n+1)/2$. If n is even, then any number between the two integers closest to $(n+1)/2$ are medians. Every integer between 1 and n is a mode.

6. Since on a majority of days, it did not rain, 0 is the median and mode of the distribution. The mean is likely to be close to 0, but depends on whether there are large outliers. These summaries would be useful in conveying that the weather is typically not rainy, but the mean would be very unuseful in understanding the typical weather patterns since it could mean that there is a constant light drizzle or mostly no rain.

   The mean, median, and mode conditioned on rainy days would be a more useful summary for what to expect on a rainy day, but would give no information about the overall weather patterns e.g. how often doe sit rain. Thus these pieces of information are complementary and both sets are useful.

7. Since log is increasing, maximizing a function is equivalent to maximizing its log. Thus we maximize $(a-1)\log x + (b-1)\log(1-x)$ whose derivative is $(a-1)/x - (b-1)/(1-x)$. Setting equal to zero, we find $(a-1)(1-x) = (b-1)x \implies x = (a-1)/(a+b-2)$ is the mode.

8. $f(x) = cx^2$. To find c, integrate $\int_0^1 cx^2dx = \frac{c}{3}x^3|_0^1 = c/3 \implies c = 3$. Then $F(x) = \int_0^x 3x^2dx = x^3$, so $F(x) = 1/2 \implies x = 1/\sqrt[3]{2}$.

9. (a) $P(Y \leq e^\mu) = P(e^X \leq e^\mu) = P(X \leq \mu)$ and similarly for $P(Y \geq e^\mu)$, so student A is correct.

   (b) This is not correct, since it identifies the random variable with its distribution.

   (c) The maximization statement is true, but the PDF of the log normal is not necessarily the exponential of the PDF of the normal.

10. Let m be the point about which X is symmetric. (In the sense that $f(m-x) = f(m+x)$ for any x or equivalently $f(x) = f(2m-x)$ for any x.) We will show that m is the mean, median, and mode.

First, m must be the unique maximum of f since otherwise by symmetry f will have multiple maxima and thus multiple modes. Furthermore by symmetry, $F(m-x) = 1 - F(m+x)$ for any x so that m is a median (i.e. take $x = 0$). If there is another median, then it would contradict unimodality, so we only need to show that m is the mean as well. The integrand in $E(m-X) = \int_{-\infty}^{\infty}(x-m)f(x)dx = \int_{-\infty}^{\infty} yf(m+y)dy$ is odd when considered as a function of y by symmetry of f: $-yf(m-y) = -(yf(m+y))$ so it is zero. Thus $E(m-X) = 0 \implies E(m) = m = E(X)$.

11. (a) $E(Z_j) = 1\sigma(E(X_j)-\mu) = 0$ and $Var(Z_j) = 1/\sigma^2 Var(X_j) = 1$. Thus the skewness of Z_j is $E(Z_j^3) = E(((X_j-\mu)/\sigma)^3) = \gamma$.

    $E(\overline{Z_n}) = 0$ and $Var(\overline{Z_n}) = 1/n^2 Var(Z_1 + \cdots + Z_n) = 1/n$. Thus the skewness of $\overline{Z_n}$ is $1/(n^{3/2})E((Z_1+\cdots+Z_n)^3) = n^{-3/2}E(((X_1-\mu)/\sigma + \cdots + (X_n-\mu)/\sigma)^3) = n^{3/2}E(((\overline{X_n}-\mu)/\sigma)^3) = E((n^{1/2}(\overline{X_n}-\mu)/\sigma)^3) = E(((\overline{X_n}-\mu)/(\sigma/n^{1/2}))^3)$ which is the skewness of $\overline{X_n}$. (Note that $Var(\overline{X_n}) = \sigma^2/n$ so the standard deviation is $\sigma/\sqrt{n}$.)

    (b) By (a), we compute the skewness of $\overline{Z_n}$ instead: $n^{-3/2}E((Z_1+\cdots+Z_n)^3) = n^{-3/2}(\sum_i E(Z_i^3) + \sum_{i<j} 3E(Z_i^2)E(Z_j) + 3E(Z_i)E(Z_j^2) + \sum_{i<j<k} 6E(Z_i)E(Z_j)E(Z_k)) = n^{-3/2}(n\gamma + \sum_{i<j} 0 + 0 + \sum_{i<j<k} 0 = \gamma/\sqrt{n})$.

    (c) When n is large, the skewness goes to 0, so the distribution of the sample mean is not skewed in either direction.

12. (a) We can apply the Cauchy-Schwarz inequality to the vectors $(X_1/n, \ldots, X_n/n)$ and $(1, \ldots, 1)$. $T_2 = n\sum(X_i/n)^2 \geq (\sum X_i/n)^2 = \overline{X_n}^2$. (So in this sense, Cauchy-Schwarz is like nonnegativity of variance.) Thus $P(T_2 > T_1) = 1$.

    (b) $E(\overline{X_n}^2) = 1/n^2(\sum_i E(X_i^2) + 2\sum_{i<j} E(X_i)E(X_j)) = 1/n^2(n(\sigma^2+c^2) + 2\binom{n}{2}c^2) = \sigma^2/n + c^2$ so the bias of T_1 is σ^2/n. (Alternatively, note that $\overline{X_n}$ is normal since it is a sum of iid normals.)

    $1/nE(\sum_j X_j^2) = 1/n\sum_j(Var(X_j) + E(X_j)^2) = \sigma^2 + c^2$

13. $E(e^{tX})E(e^{tY}) = (\sum_{j=1}^{6} e^{tj}/6)^2$

14. $E(e^{tX}) = \prod_i E(e^{tU_i}) = \prod_i \int_0^1 e^{tx}dx = ((e^t-1)/t)^{60}$

15. (a) $E(e^{tW}) = E(e^{tX^2})^2$. For the MGF of X^2, note that $E(e^{tX^2}) = \int_{-\infty}^{\infty} e^{tz^2}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{z^2(t-1/2)}dz = \frac{1}{\sqrt{1-2t}}$. This converges for $t < 1/2$ so it will give a valid MGF. Therefore the MGF of W is $1/(1-2t)$.

    (b) This is an exponential distribution with parameter $\lambda = 1/2$: $\int_0^{\infty}\lambda e^{(t-\lambda)x}dx = \frac{\lambda}{\lambda-t} = \frac{1}{1-t/\lambda}$.

16. For any positive a, the skewness of aX is $E(((aX-a\mu)/(a\sigma))^3) = E(((X-\mu)/\sigma)^3)$ which is equal to the skewness of X. In particular, the skewness of X is the same as the skewness of $\lambda X \sim Expo(1)$, which is $E((X-1)^3) = E(X^3 - 3X^2 + 3X - 1) = 6 - 3*2 + 3 - 1 = 2$.

17. (a) $\overline{X_n}$ has mean μ and variance σ^2/n, so Z_n is the standardization by definition.

    (b) $Z_n = \sum_i (X_i - \mu)/\sqrt{n}\sigma$ so $E(e^{tZ_n}) = \prod_i E(e^{t(X_i-\mu)/\sqrt{n}\sigma}) = e^{-\mu t\sqrt{n}/\sigma} M(t/\sqrt{n}\sigma)^n$

18. $E(e^{tX}) = \sum_k e^{tk} q^k p = p/(1-qe^t)$. The derivative is $p(qe^t)(1-qe^t)^{-2}$ which is $pq/(1-q)^2 = q/p$ when $t=0$ giving the mean.

    The second moment is $p(qe^t)(1-qe^t)^{-2} + 2p(qe^t)^2(1-qe^t)^{-3}$ which is $pq/(1-q)^2 + 2pq^2/(1-q)^3 = q/p + 2q^2/p^2$ when $t=0$ giving the second moment. Then the difference is $q/p + q^2/p^2 = (q/p)(1+q/p) = q/p^2$.

19. $E(e^{2X+Y}) = e^{\lambda(e^{2t}-1)} e^{\lambda(e^t-1)} = e^{\lambda(e^{2t}+e^t-2)}$ is not the MGF of a Poisson distribution.

20. $M(t) = e^{\lambda(e^t-1)} \implies g(t) = \lambda(e^t-1) = \sum_{j\geq 1} \lambda t^j/j!$ so all cumulants are λ.

21. $E(e^{tX_n}) = \prod_i E(e^{tI_j})$ where $X = \sum_j I_j$ is a sum of independent Bernoulli rv's. Then $E(e^{tI_j}) = pe^t + q$ so $E(e^{tX_n}) = (\lambda e^t/n + 1 - \lambda/n)^n = (1 + 1/n(\lambda(e^t-1)))^n \xrightarrow{n\to\infty} e^{\lambda(e^t-1)}$ which is the Poisson MGF.

22. (a) $E(e^{tX}) = \prod E(e^{tI_j}) = \prod_j (p_j e^t + q_j)$

    (b) $(p_j e^t + 1 - p_j) = (1 + p_j(e^t - 1)) \approx e^{p_j(e^t-1)}$ for each j, so the MGF becomes $\approx \prod_j e^{p_j(e^t-1)} = e^{\lambda(e^t-1)}$. This is the Poisson MGF and this makes sense intuitively since this is exactly the situation where the Poisson approximation should be valid.

23. $E(e^{t(U_1+U_2)}) = E(e^{tU_1})E(e^{tU_2}) = (e^t-1)^2/t^2$.

    $E(e^{tT}) = \int_0^1 e^{tx} x dx + \int_1^2 e^{tx}(2-x)dx = (e^t(t-1)+1)/t^2 + e^t(-t+e^t-1)/t^2 = (e^t-1)^2/t^2$ so these random variables have the same MGF.

24. $E(e^{tL}) = \int_{-\infty}^{\infty} e^{tx} e^{-|x|}/2dx = \int_{-\infty}^{0} e^{tx} e^{x}/2dx + \int_0^{\infty} e^{tx} e^{-x}/2dx = 1/21/(t+1) - 1/21/(t-1) = 1/(t+1)(t-1)$.

    $E(e^{t(X-Y)}) = E(e^{tX})E(e^{-tY}) = 1/(1-t)1/(1+t)$ so they have the same MGF and have L is Laplace.

25. $E(e^{tY}) = \int_0^{\infty} e^{ty} f(y)dy = \int_{-\infty}^{\infty} e^{t|z|}\varphi(z)dz$.

26. (a) $E(\Phi(Z)e^Z) = \int_{-\infty}^{\infty} \Phi(z)e^z\varphi(z)dz$.

    (b) $E(\Phi(Z)) = .5$ since $\varphi(Z)$ is uniform. $E(e^Z) = e^{1/2}$ since this is the MGF evaluated at $t=1$.