STAT/MA 41600
In-Class Problem Set #41: December 3, 2014
Solutions by Mark Daniel Ward

1a. Let X be the waiting time. Then, by Markov's inequality, we have $P(X > 10) \leq 8/10$.

1b. We have $P(6.5 < X < 9.5) = P(|X - 8| < 1.5)$. If we have $k\sigma_X = 1.5$, then (since $\sigma_X = 1.5$), we have $k = 1$. Thus, by Chebyshev's Inequality, we have $P(6.5 < X < 9.5) = P(|X - 8| < (1)\sigma_X) \geq 1 - 1/1^2 = 0$. So our inequality is trivial in this case. *Indeed, we get a bound of "0" like this, whenever we use Chebyshev's inequality to find the probability that a random variable is within 1 standard deviation away from its mean.*

2. Let X denote the waiting time. Then $P(2 < X < 3) = P(|X - 2.5| < 0.5)$. If we have $k\sigma_X = 0.5$, then (since $\sigma_X = 1.5$), we have $k = 1/3$. Thus, by Chebyshev's Inequality, we have $P(2 < X < 3) = P(|X - 2.5| < (1/3)\sigma_X) \geq 1 - 1/(1/3)^2 = -8$. So our inequality is even worse than trivial in this case! We already knew that $P(2 < X < 3) \geq 0$, so this application of Chebyshev's inequality is completely true but it is not too helpful. (All probabilities are already known to be nonnegative.) *Indeed, we get a negative lower bound like this, whenever we use Chebyshev's inequality to find the probability that a random variable is less than 1 standard deviation away from its mean.*

3. Let X denote the time spent sleeping. Then $P(X < 5 \text{ or } X > 7) = P(|X - 6| > 1)$. If we have $k\sigma_X = 1$, then (since $\sigma_X = 1.3$), we have $k = 1/(1.3)$. Thus, by Chebyshev's Inequality, we have $P(X < 5 \text{ or } X > 7) = P(|X - 6| > (1/(1.3))(1.3)) \leq (1.3)^2 = 1.69$. So our inequality is even worse than trivial in this case! We already knew that $P(X < 5 \text{ or } X > 7) \leq 1$, so this application of Chebyshev's inequality is (again) completely true but it is not too helpful. (All probabilities are already known to be at most 1.) *Indeed, we get a bound greater than 1 like this, whenever we use Chebyshev's inequality to find the probability that a random variable is more than 1 standard deviation away from its mean.*

4a. The sum $X_1 + \cdots + X_{250}$ is a negative binomial random variable with $r = 250$ and $p = 1/3$, so $P(730 \leq X_1 + \cdots + X_{250} \leq 770) = \sum_{x=730}^{770} \binom{x-1}{249} (1/3)^{250} (2/3)^{x-250}$. (FYI, you probably cannot calculate this on your calculator, but if we use a software program, we can calculate the exact value as $0.40340111677\ldots$.)

4b. Since $X_1 + \cdots + X_{250}$ has expected value $\frac{250}{1/3} = 750$ and variance $\frac{(2/3)(250)}{(1/3)^2} = 1500$, we can make an approximation of $X_1 + \cdots + X_{250}$ to a Normal random variable (using continuity correction since $X_1, \ldots, X_{250}$ are discrete valued), and we get $P(730 \leq X_1 + \cdots + X_{250} \leq 770) = P(729.5 \leq X_1 + \cdots + X_{250} \leq 770.5) = P(\frac{729.5-750}{\sqrt{1500}} \leq \frac{X_1+\cdots+X_{250}-750}{\sqrt{1500}} \leq \frac{770.5-750}{\sqrt{1500}}) \approx P(-0.53 \leq Z \leq 0.53) = P(Z \leq 0.53) - P(Z \leq -0.53) = P(Z \leq 0.53) - P(Z \geq 0.53) = P(Z \leq 0.53) - (1 - P(Z \leq 0.53)) = 0.7019 - (1 - 0.7019) = 0.4038$. That is a pretty accurate value. It agrees with the actual value (from part **a**) in the first three decimal places.

5a. *Method #1:* Since the joint density is constant, we can make the calculation using areas. The area of the entire region where X and Y are defined is 9. The area of the region where $Y < 1$ and $X < 1$ is $(1)(1) = 1$. So $P(Y < 1 \text{ and } X < 1) = 1/9$. The area of the region where $X < 1$ is 5, so $P(X < 1) = 5/9$. Thus $P(Y < 1 \mid X < 1) = \frac{P(Y<1 \text{ and } X<1)}{P(X<1)} = \frac{1/9}{5/9} = 1/5$.

Method #2: The area of the triangle is 9, so the joint density of X and Y is $f_{X,Y}(x,y) = 1/9$. Therefore, we get $P(Y < 1 \text{ and } X < 1) = \int_0^1 \int_0^1 1/9\,dy\,dx = 1/9$ and $P(X < 1) = \int_0^1 \int_0^{6-2x} 1/9\,dy\,dx = 5/9$. So we conclude $P(Y < 1 \mid X < 1) = \frac{P(Y<1 \text{ and } X<1)}{P(X<1)} = \frac{1/9}{5/9} = 1/5$.

5b. We have $f_{Y|X}(y \mid 1) = \frac{f_{X,Y}(1,y)}{f_X(1)}$. The numerator is $1/9$ for $0 \le y \le 6-2x$, and 0 otherwise. We then calculate $f_X(x) = \int_0^{6-2x} 1/9\,dy = (6-2x)/9$ for $0 \le x \le 3$, so the denominator is $f_X(1) = (6-(2)(1))/9 = 4/9$. So the conditional density is $f_{Y|X}(y \mid 1) = \frac{1/9}{4/9} = 1/4$ for $0 \le y \le 6-(2)(1) = 4$. Finally, we conclude $P(Y < 1 \mid X = 1) = \int_0^1 f_{Y|X}(y \mid 1)\,dy = \int_0^1 1/4\,dy = 1/4$.
5c. We have $\mathbb{E}(Y \mid X = 1) = \int_0^4 (y) f_{Y|X}(y \mid 1)\,dy = \int_0^4 (y)(1/4) = 2$.

6. We define $X_1 + \cdots + X_7$ where X_j indicates if Gwyneth's jth pick is a blues CD, or $X_j = 0$ otherwise. Similarly, we define $Y_1 + \cdots + Y_7$ where Y_k indicates if Josephine's kth pick is a blues CD. We have $\mathrm{Cov}(X,Y) = \mathrm{Cov}(X_1 + \cdots + X_7, Y_1 + \cdots + Y_7) = \sum_{j=1}^7 \sum_{k=1}^7 \mathrm{Cov}(X_j, Y_k)$. For any j with $1 \le j \le 7$ and k with $1 \le k \le 7$, we have $\mathrm{Cov}(X_j, Y_k) = \mathbb{E}(X_j Y_k) - \mathbb{E}(X_j)\mathbb{E}(Y_k) = (5/20)(4/19) - (5/20)(5/20) = -3/304$. Thus $\mathrm{Cov}(X,Y) = (49)(-3/304) = -147/304$.