WINTER 2017/18 TERM 2 STAT 302: ASSIGNMENT 4
Due: 2pm on Tuesday April 10, 2018

Question 1: Let Y denote the number of bacteria per cubic centimeter in a particular liquid and suppose $Y \sim Poisson(\lambda)$. Further, assume that λ varies with location and has a Gamma distribution with parameters α and β.

(a) What is the expected number of bacteria per cubic centimeter?

Since $Y|\Lambda \sim Poisson(\Lambda)$, the expected number of bacteria $\mathbb{E}_{Y|\Lambda}(Y|\Lambda = \lambda) = \lambda$.
Since $\Lambda \sim Gamma(\alpha, \beta)$, the expected value $\mathbb{E}_\Lambda(\Lambda) = \frac{\alpha}{\beta}$
By double expectation formula, the expected number of bacteria per cubic centimeter $\mathbb{E}(Y) = \mathbb{E}_\Lambda[\mathbb{E}_{Y|\Lambda}(Y|\Lambda = \lambda)] = \mathbb{E}_\Lambda(\Lambda) = \frac{\alpha}{\beta}$

(b) What is the standard deviation of the number of bacteria per cubic centimeter?

$Y|\Lambda \sim Poisson(\Lambda), so\ \mathbb{E}_{Y|\Lambda}(Y|\Lambda = \lambda) = \lambda, \mathrm{Var}_{Y|\Lambda}(Y|\Lambda = \lambda) = \lambda.$
$\Lambda \sim Gamma(\alpha, \beta), so\ \mathbb{E}_\Lambda(\Lambda) = \frac{\alpha}{\beta}, \mathrm{Var}_\Lambda(\Lambda) = \frac{\alpha}{\beta^2}.$

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathbb{E}_\Lambda[\mathrm{Var}(Y|\Lambda)] + \mathrm{Var}_\Lambda[\mathbb{E}(Y|\Lambda)] \\
&= \mathbb{E}_\Lambda[\Lambda] + \mathrm{Var}_\Lambda[\Lambda] \\
&= \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2} \\
&= \frac{\alpha\beta + \alpha}{\beta^2}
\end{aligned}$$

Therefore, the standard deviation of the number of bacteria per cubic centimeter $\mathrm{SD}(Y) = \sqrt{\mathrm{Var}(Y)} = \sqrt{\frac{\alpha\beta+\alpha}{\beta^2}} = \frac{\sqrt{\alpha\beta+\alpha}}{\beta}$.

(c) What is the probability that $Y \geqslant 1$?

$$joint\ pdf = f(y, \lambda) = f_{Y|\Lambda}(y|\lambda) \cdot f_\Lambda(\lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \cdot \frac{\beta e^{-\beta\lambda}(\beta\lambda)^{\alpha-1}}{\Gamma(\alpha)},\ y = 0, 1, 2, ...,\ \lambda \geqslant 0$$

$$f_{Y|\Lambda}(y = 0, \lambda) = e^{-\lambda}\frac{\beta e^{-\beta\lambda}(\beta\lambda)^{\alpha-1}}{\Gamma(\alpha)},\ \lambda \geqslant 0$$

$$\begin{aligned}
Pr(Y = 0) &= \int_0^\infty e^{-\lambda}\frac{\beta e^{-\beta\lambda}(\beta\lambda)^{\alpha-1}}{\Gamma(\alpha)} d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty e^{-\lambda(1+\beta)}\lambda^{\alpha-1} d\lambda, && pull\ out\ constants \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty e^{-y}(\frac{y}{1+\beta})^{\alpha-1}\frac{1}{1+\beta} dy, && substitute\ y = \lambda(1+\beta) \\
&= (\frac{\beta}{1+\beta})^\alpha \frac{1}{\Gamma(\alpha)}\int_0^\infty e^{-y}y^{\alpha-1} dy, && pull\ out\ constants \\
&= (\frac{\beta}{1+\beta})^\alpha \frac{1}{\Gamma(\alpha)}\Gamma(\alpha), && by\ definition\ of\ \Gamma(\alpha) \\
&= (\frac{\beta}{1+\beta})^\alpha
\end{aligned}$$

Therefore, the probability that $Y \geqslant 1$ is

$$Pr(Y \geqslant 1) = 1 - Pr(Y = 0) = 1 - (\frac{\beta}{1+\beta})^{\alpha}$$

Question 2: Let X_1 and X_2 be given by the joint probability density function:

$$f(x_1, x_2) = \begin{cases} 6(1-x_2), & 0 \leqslant x_1 \leqslant x_2 \leqslant 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find $\mathrm{Cov}(X_1, X_2)$.

$$f_{X_1}(x_1) = \int_{x_1}^{1} 6(1-x_2)dx_2 = 6(x_2 - \frac{1}{2}x_2^2)_{x_1}^{1} = 3(x_1-1)^2,\ 0 \leqslant x_1 \leqslant 1$$
$$\mathbb{E}(X_1) = \int_0^1 3x_1(x_1-1)^2 dx_1 = 3\int_0^1 x_1^3 - 2x_1^2 + x_1 dx_1 = \frac{1}{4}$$
$$f_{X_2}(x_2) = \int_0^{x_2} 6(1-x_2)dx_1 = 6(x_2 - x_2^2),\ 0 \leqslant x_2 \leqslant 1$$
$$\mathbb{E}(X_2) = \int_0^1 6x_2(x_2 - x_2^2)dx_2 = 6(\frac{1}{3}x_2^3 - \frac{1}{4}x_2^4)_0^1 = \frac{1}{2}$$
$$\mathbb{E}(X_1X_2) = \int_0^1\int_0^{x_2} x_1x_2 6(1-x_2)dx_1dx_2 = 6\int_0^1 [\frac{1}{2}x_1^2(x_2 - x_2^2)]_0^{x_2} dx_2 = 3\int_0^1 x_2^3 - x_2^4 dx_2 = \frac{3}{20}$$
$$Cov(X_1, X_2) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = \frac{3}{20} - \frac{1}{4} \times \frac{1}{2} = \frac{1}{40}$$

(b) Are X_1 and X_2 independent?

X_1 and X_2 are **not** independent. We can prove it by contradiction. If X_1 and X_2 are independent, then $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2)$. From (a) we know that $\mathbb{E}(X_1X_2) = \frac{3}{20}$, and $\mathbb{E}(X_1) \cdot \mathbb{E}(X_2) = \frac{1}{8}$, so $\mathbb{E}(X_1X_2) \neq \mathbb{E}(X_1) \cdot \mathbb{E}(X_2)$. Contradiction! Therefore X_1 and X_2 are not independent.

(c) Compute $\mathrm{Var}(3X_1 - X_2)$.

$$\mathbb{E}(X_1^2) = \int_0^1 3x_1^2(x_1-1)^2dx_1 = 3\int_0^1 x_1^4 - 2x_1^3 + x_1^2 dx_1 = 3(\frac{1}{5}x_1^5 - \frac{1}{2}x_1^4 + \frac{1}{3}x_1^3)_0^1 = \frac{1}{10}$$
$$\mathrm{Var}(X_1) = \mathbb{E}(X_1^2) - [\mathbb{E}(X_1)]^2 = \frac{1}{10} - (\frac{1}{4})^2 = \frac{3}{80}$$
$$\mathbb{E}(X_2^2) = \int_0^1 6x_2^2(x_2 - x_2^2)dx_2 = 6(\frac{1}{4}x_2^4 - \frac{1}{5}x_2^5)_0^1 = \frac{3}{10}$$
$$\mathrm{Var}(X_2) = \mathbb{E}(X_2^2) - [\mathbb{E}(X_2)]^2 = \frac{3}{10} - (\frac{1}{2})^2 = \frac{1}{20}$$
$$\begin{aligned} &\mathrm{Var}(3X_1 - X_2) \\ &= \mathrm{Var}(3X_1) + \mathrm{Var}(-X_2) + 2Cov(3X_1, -X_2) \\ &= 9\mathrm{Var}(X_1) + \mathrm{Var}(X_2) - 6Cov(X_1, X_2) \\ &= 9 \times \frac{3}{80} + \frac{1}{20} - 6 \times \frac{1}{40} \\ &= \frac{19}{80} \end{aligned}$$

Question 3: Recent Statistics Canada data indicate that among Canadians who do not develop diabetes or hypertension in their lifetimes, the average life expectancy for females is 85.5 years while the average life expectancy for males is 81.0 years. Females also outnumber males in Canada by a slim margin of 50.4% to 49.6% of the total population. Suppose we randomly select a Canadian with negligible chance of developing diabetes or hypertension (i.e. healthy diet, active, non-smoker, no family history of afflictions).

(a) Find a bound on the probability that this person will live to be more than 100 years old.

Let F denote the life expectancy of a female Canadian. $\mathbb{E}(F) = 85.5$
Let M denote the life expectancy of a male Canadian. $\mathbb{E}(M) = 81.0$
Let X denote the life expectancy of a Canadian.X = 50.4% F + 49.6% M = 0.504F + 0.496M.
$\mathbb{E}(X) = \mathbb{E}(0.504F) + \mathbb{E}(0.496M) = 0.504 \times 85.5 + 0.496 \times 81.0 = 83.268$
According to Markov's Inequality,

$$Pr(X \geqslant 100) \leqslant \frac{\mathbb{E}(X)}{100} = \frac{83.268}{100} = 0.83268$$

Therefore, the probability of this person to live more than 100 years is less than or equal to 83.268%.

(b) Statistics Canada also reports that the standard deviation of life expectancy in individuals who do not develop diabetes or hypertension is about 7 years for women, and about 8 years for men. Using this new information, find a better bound on the probability that our randomly selected individual will live to be more than 100 years old.

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(0.504F + 0.496M)\\ &= 0.504^2\mathrm{Var}(F) + 0.496^2\mathrm{Var}(M),\ \textit{M and F are independent}\\ &= 0.504^2 \times 7^2 + 0.496^2 \times 8^2\\ &= 28.191808\end{aligned}$$

Therefore, $\mu = 83.268, \sigma^2 = 28.191808, k = 100 - 83.268 = 16.732$
According to Chebyshev's Inequality:

$$Pr(|X - \mu| \geqslant k) \leqslant \frac{\sigma^2}{k^2}$$

$$\textit{plug in } \mu, k, \sigma^2,\ Pr(|X - 83.268| \geqslant 16.732) \leqslant \frac{28.191808}{16.732^2}$$

$$Pr(X > 100) \leqslant Pr(X \leqslant 66.536 \textit{ or } X \geqslant 100) \leqslant 0.1007$$

Therefore, the probability of the randomly selected individual to live more than 100 years is less than or equal to 10.07%.

(c) Now suppose we select 50 men at random who have a negligible chance of developing diabetes or hypertension. What is the chance that their average life expectancy will be less than 80 years? How does this compare to the chance of the same event if we were to select 50 healthy women instead?

Let $M_1, M_2, ...$ be a sequence of independent and identically distributed random variables, each having $\mathbb{E}(M_i) = \mu_M = 81.0$, and $\mathrm{Var}(M_i) = \sigma_M^2 = 8^2$. According to Central Limit

Theorem,

$$
\begin{aligned}
Z_n &= \frac{\sqrt{n}(\bar{X}_n - \mu_M)}{\sigma_M} \\
z &= \frac{\sqrt{n}(80 - \mu_M)}{\sigma_M} = \frac{\sqrt{50}(80 - 81.0)}{8} = -\frac{5\sqrt{2}}{8} = -0.8839 \\
Pr(Z_n &\leqslant z) = Pr(Z_n \leqslant -0.8839) = 0.1884
\end{aligned}
$$

Therefore, the probability of these sample men's average life expectancy to be less than 80 years is 0.1884.

Let $F_1, F_2, ...$ be a sequence of independent and identically distributed random variables, each having $\mathbb{E}(F_i) = \mu_F = 85.5$, and $\text{Var}(F_i) = \sigma_F^2 = 7^2$. According to Central Limit Theorem,

$$
\begin{aligned}
Z_n &= \frac{\sqrt{n}(\bar{X}_n - \mu_F)}{\sigma_F} \\
z &= \frac{\sqrt{n}(80 - \mu_F)}{\sigma_F} = \frac{\sqrt{50}(80 - 85.5)}{7} = -\frac{55\sqrt{2}}{14} = -5.55584 \\
Pr(Z_n &\leqslant z) = Pr(Z_n \leqslant -5.55584) \approx 0
\end{aligned}
$$

Therefore, the probability of these sample women average life expectancy to be less than 80 years is almost 0.