HOMEWORK 01

PSTAT 120B: Mathematical Statistics, I
Summer Session A, 2024 with Instructor: Ethan P. Marzban

---

1. (**PSTAT 120A Review**) For a random variable X and constant a and b, show that

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$

Solution:

$$\begin{aligned} \mathrm{Var}(aX+b) &= \mathbb{E}[(aX+b)^2] - (\mathbb{E}[aX+b])^2 \\ &= \mathbb{E}[a^2X^2 + 2abX + b^2] - (a\mathbb{E}[X] + b)^2 \\ &= a^2\mathbb{E}[X^2] + \cancel{2ab\mathbb{E}[X]} + \cancel{b^2} - a^2(\mathbb{E}[X])^2 - \cancel{2ab\mathbb{E}[X]} - \cancel{b^2} \\ &= a^2\left\{\mathbb{E}[X^2] - (\mathbb{E}[X])^2\right\} \\ &= a^2\mathrm{Var}(X) \qquad \blacksquare \end{aligned}$$

2. (**Modified from #5.36**) Let Y_1 and Y_2 denote the proportions of time (out of one workday) during which employees I and II, respectively, perform their assigned tasks. The joint relative frequency behavior of Y_1 and Y_2 is modeled by the density function

$$f_{Y_1,Y_2}(y_1,y_2) = (y_1+y_2)\cdot \mathbb{1}_{\{0\le y_1 \le 1,\ 0 \le y_2 \le 1\}}$$

(a) Verify that this is a valid joint density function.

Solution: Recall that a function need only satisfy two conditions in order to be a valid density: nonnegativity, and integrating to unity. Nonnegativity is fairly trivial; for any $y_1 \in [0,1]$ and $y_2 \in [0,1]$ we have that both y_1 and y_2 are nonnegative, and hence their sum will also be nonnegative- thus, $f_{Y_1,Y_2}(y_1,y_2) \ge 0$ for every $(y_1,y_2) \in \mathbb{R}$.

To check integration to unity, we compute

$$\begin{aligned} \iint_{\mathbb{R}^2} f_{Y_1,Y_2}(y_1,y_2)\,\mathrm{d}A &= \int_0^1 \int_0^1 (y_1+y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 \\ &= \int_0^1 \left[\frac{1}{2}y_1^2 + y_1y_2\right]_{y_1=0}^{y_1=1} \mathrm{d}y_2 \\ &= \int_0^1 \left(\frac{1}{2} + y_2\right) \mathrm{d}y_2 \\ &= \left[\frac{1}{2}y_2 + \frac{1}{2}y_2^2\right]_{y_2=0}^{y_2=1} = \frac{1}{2} + \frac{1}{2} = 1 \checkmark \end{aligned}$$

Hence, we can conclude that $f_{Y_1,Y_2}(y_1,y_2)$ is a valid joint density function.

(b) Find the marginal density functions for Y_1 and Y_2.

Solution:

$$\begin{aligned}
f_{Y_1}(y_1) &= \int_{\mathbb{R}} f_{Y_1,Y_2}(y_1,y_2)\,\mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} (y_1+y_2)\cdot \mathbb{1}_{\{0\le y_1\le 1\}}\cdot \mathbb{1}_{\{0\le y_2\le 1\}}\,\mathrm{d}y_2 \\
&= \mathbb{1}_{\{0\le y_1\le 1\}}\cdot \int_0^1 (y_1+y_2)\,\mathrm{d}y_2 \\
&= \mathbb{1}_{\{0\le y_1\le 1\}}\cdot \left[y_1y_2+\frac{1}{2}y_2^2\right]_{y_2=0}^{y_2=1} \\
&= \boxed{\left(\frac{1}{2}+y_1\right)\cdot \mathbb{1}_{\{0\le y_1\le 1\}}} \\
f_{Y_2}(y_2) &= \int_{\mathbb{R}} f_{Y_1,Y_2}(y_1,y_2)\,\mathrm{d}y_1 \\
&= \int_{-\infty}^{\infty} (y_1+y_2)\cdot \mathbb{1}_{\{0\le y_1\le 1\}}\cdot \mathbb{1}_{\{0\le y_2\le 1\}}\,\mathrm{d}y_1 \\
&= \mathbb{1}_{\{0\le y_2\le 1\}}\cdot \int_0^1 (y_1+y_2)\,\mathrm{d}y_1 \\
&= \mathbb{1}_{\{0\le y_2\le 1\}}\cdot \left[\frac{1}{2}y_1^2+y_1y_2\right]_{y_1=0}^{y_1=1} \\
&= \boxed{\left(\frac{1}{2}+y_2\right)\cdot \mathbb{1}_{\{0\le y_2\le 1\}}}
\end{aligned}$$

(We could have also surmised the density of $f_{Y_2}(y_2)$ through symmetry.)

(c) Are Y_1 and Y_2 independent? Why or why not? Be careful about your justification!

Solution: $Y_1 \perp Y_2$ only if their joint density factors as a product of their marginals. From our answers to part (b) above, we find

$$\begin{aligned}
f_{Y_1}(y_1)\cdot f_{Y_2}(y_2) &= \left(\frac{1}{2}+y_1\right)\cdot \mathbb{1}_{\{0\le y_1\le 1\}}\cdot \left(\frac{1}{2}+y_2\right)\cdot \mathbb{1}_{\{0\le y_2\le 1\}} \\
&= \left(\frac{1}{4}+\frac{1}{2}y_1+\frac{1}{2}y_2+y_1y_2\right)\cdot \mathbb{1}_{\{0\le y_1\le 1,\,0\le y_2\le 1\}} \\
&\neq (y_1+y_2)\cdot \mathbb{1}_{\{0\le y_1\le 1,\,0\le y_2\le 1\}} = f_{Y_1,Y_2}(y_1,y_2)
\end{aligned}$$

So, since $f_{Y_1,Y_2}(y_1,y_2)\neq f_{Y_1}(y_1)\cdot f_{Y_2}(y_2)$, we have that Y_1 and Y_2 are not independent.

(d) Find $\mathbb{P}(Y_1 \ge 1/2 \mid Y_2 \ge 1/2)$.

Solution: By the definition of conditional probability,

$$\mathbb{P}(Y_1 \geq 1/2 \mid Y_2 \geq 1/2) = \frac{\mathbb{P}(\{Y_1 \geq 1/2\} \cap \{Y_2 \geq 1/2\}}{\mathbb{P}(Y_2 \geq 1/2)}$$

The numerator can be computed by double-integrating the joint density, and the denominator can be found by integrating the marginal density of Y_2 that we derived in part (b) above.

$$\begin{aligned}
\mathbb{P}(Y_1 \geq 1/2,\ Y_2 \geq 1/2) &= \int_{1/2}^{1} \int_{1/2}^{1} (y_1 + y_2)\ \mathsf{d}y_1\ \mathsf{d}y_2 \\
&= \int_{1/2}^{1} \left[\frac{1}{2} y_1^2 + y_1 y_2 \right]_{y_1 = 1/2}^{y_1 = 1} \mathsf{d}y_2 \\
&= \int_{1/2}^{1} \left(\frac{1}{2} + y_2 - \frac{1}{8} - \frac{1}{2} y_2 \right) \mathsf{d}y_2 \\
&= \int_{1/2}^{1} \left(\frac{3}{8} + \frac{1}{2} y_2 \right) \mathsf{d}y_2 \\
&= \left[\frac{3}{8} y_2 + \frac{1}{4} y_2^2 \right]_{y_2 = 1/2}^{y_2 = 1} \\
&= \frac{3}{8} + \frac{1}{4} - \frac{3}{16} - \frac{1}{16} = \frac{3}{8} \\
\mathbb{P}(Y_2 \geq 1/2) &= \int_{1/2}^{\infty} f_{Y_2}(y_2)\ \mathsf{d}y_2 \\
&= \int_{1/2}^{1} \left(\frac{1}{2} + y_2 \right) \mathsf{d}y_2 \\
&= \left[\frac{1}{2} y_2 + \frac{1}{2} y_2^2 \right]_{y_2 = 1/2}^{y_2 = 1} \\
&= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} - \frac{1}{8} = \frac{5}{8}
\end{aligned}$$

Hence, putting everything together,

$$\mathbb{P}(Y_1 \geq 1/2 \mid Y_2 \geq 1/2) = \frac{\mathbb{P}(\{Y_1 \geq 1/2\} \cap \{Y_2 \geq 1/2\}}{\mathbb{P}(Y_2 \geq 1/2)} = \frac{3/8}{5/8} = \frac{3}{5}$$

(e) If employee II spends exactly 50% of the day working on assigned duties, find the probability that employee I spends more than 75% of the day working on similar duties.

Solution: This part is asking us to compute $\mathbb{P}(Y_1 \geq 3/4 \mid Y_2 = 1/2)$, which requires us to first find the conditional density $f_{Y_1|Y_1}(y_1 \mid y_2)$ [since the event we are conditioning on, $\{Y_2 = 1/2\}$, has zero probability]. We do so by using the definition of conditional densities:

$$f_{Y_1|Y_2}(y_1 \mid y_2) = \frac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_2}(y_2)}$$

$$= \frac{(y_1 + y_2) \cdot \mathbb{1}_{\{0 \le y_1 \le 1\}} \cdot \cancel{\mathbb{1}_{\{0 \le y_2 \le 1\}}}}{\left(\frac{1}{2} + y_2\right) \cdot \cancel{\mathbb{1}_{\{0 \le y_2 \le 1\}}}}$$
$$= \frac{y_1 + y_2}{\frac{1}{2} + y_2} \cdot \mathbb{1}_{\{0 \le y_1 \le 1\}}$$

This tells us that, plugging in $y_2 = 1/2$,

$$f_{Y_1 | Y_2}(y_1 \mid 1/2) = \frac{y_1 + 1/2}{\frac{1}{2} + \frac{1}{2}} \cdot \mathbb{1}_{\{0 \le y_1 \le 1\}} = \left(\frac{1}{2} + y_1\right) \cdot \mathbb{1}_{\{0 \le y_1 \le 1\}}$$

and so

$$\begin{aligned} \mathbb{P}(Y_1 \ge 3/4 \mid Y_2 = 1/2) &= \int_{3/4}^{\infty} f_{Y_1|Y_2}(y_1 \mid 1/2) \, \mathrm{d}y_1 \\ &= \int_{3/4}^{1} \left(\frac{1}{2} + y_1\right) \, \mathrm{d}y_1 \\ &= \left[\frac{1}{2} y_1 + \frac{1}{2} y_1^2\right]_{y_1 = 3/4}^{y_1 = 1} \\ &= \frac{1}{2} + \frac{1}{2} - \frac{3}{8} - \frac{9}{32} = \boxed{\frac{11}{32}} \end{aligned}$$

which is equivalent to 34.375%.

3. (**Modified from #5.157**) A forester studying diseased pine trees models the number of diseased trees per acre, Y , as a Poisson random variable with mean λ. However, λ changes from area to area, and its random behavior is modeled by a gamma distribution. That is, for some integer α and a positive constant $\beta > 0$,

$$f(\lambda) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \lambda^{\alpha - 1} e^{-\lambda/\beta} \cdot \mathbb{1}_{\{\lambda \ge 0\}}$$

Find the <u>unconditional</u> probability distribution of Y. Because α is assumed to be an integer, you should be able to recognize this distribution by name - include any/all relevant parameter(s)! **Hint:** When integrating/summing, try to multiply and divide by constants to get a density function inside the integral/sum. This will then avoid you having to perform any direct integration/summation!

Solution: Let Λ denote the random variable corresponding to the rate of diseased trees. From the problem statement,

$$\begin{aligned} (Y \mid \Lambda = \lambda) &\sim \text{Pois}(\lambda) \\ \Lambda &\sim \text{Gamma}(\alpha, \beta) \end{aligned}$$

Hence, we have access to $p_{Y|\Lambda}(y \mid \lambda)$ and $f_{\Lambda}(\lambda)$, and we seek $p_Y(y)$ [note that Y will be discrete]. The trick is to use the continuous-case formula from slide 31 of the Topic01 Slides (we use the continuous-case since the random variable we are conditioning on, Λ, is continuous):

$$p_Y(y) = \int_{\mathbb{R}} p_{Y|\Lambda}(\lambda) \cdot f_{\Lambda}(\lambda) \, \mathrm{d}\lambda$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} e^{-\lambda} \cdot \frac{\lambda^y}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} \cdot \mathbb{1}_{\{\lambda \geq 0\}} \, \mathrm{d}\lambda \\
&= \frac{1}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \int_0^\infty \lambda^{(\alpha+y)-1} \cdot e^{-\lambda\left(1+\frac{1}{\beta}\right)} \, \mathrm{d}\lambda
\end{aligned}
$$

There are a couple of ways to solve this particular integral: one is to make a $u-$substitution and then manipulate terms to turn the integral into a Gamma function, and the other is to multiply and divide by appropriate constants to transform the integrand into the density of a Gamma *distribution*. I'll start by demonstrating the latter:

$$
\begin{aligned}
p_Y(y) &= \frac{1}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \int_0^\infty \lambda^{(\alpha+y)-1} \cdot e^{-\lambda \Big/ \left[\frac{1}{\left(1+\frac{1}{\beta}\right)}\right]} \, \mathrm{d}\lambda \\
&= \frac{1}{y!} \cdot \frac{\Gamma(\alpha+y) \cdot \left[\frac{1}{\left(1+\frac{1}{\beta}\right)}\right]^{\alpha+y}}{\Gamma(\alpha) \cdot \beta^\alpha} \cdot \int_0^\infty \frac{1}{\Gamma(\alpha+y) \cdot \left[\frac{1}{\left(1+\frac{1}{\beta}\right)}\right]^{\alpha+y}} \cdot \lambda^{(\alpha+y)-1} \cdot e^{-\lambda \Big/ \left[\frac{1}{\left(1+\frac{1}{\beta}\right)}\right]} \, \mathrm{d}\lambda
\end{aligned}
$$

The integrand is now the density of a Gamma$(\alpha + y \ , \ 1/(1+1/\beta))$ distribution. Since we are integrating this density over its entire support, the integral must be unity:

$$
\begin{aligned}
p_Y(y) &= \frac{1}{y!} \cdot \frac{\Gamma(\alpha+y) \cdot \left[\frac{1}{\left(1+\frac{1}{\beta}\right)}\right]^{\alpha+y}}{\Gamma(\alpha) \cdot \beta^\alpha} \\
&= \frac{1}{y!} \cdot \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)} \cdot \frac{1}{\beta^\alpha} \cdot \left[\frac{1}{\left(\frac{\beta+1}{\beta}\right)}\right]^{\alpha+y} \\
&= \frac{1}{y!} \cdot \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)} \cdot \frac{1}{\cancel{\beta^\alpha}} \cdot \frac{\beta^{\cancel{\alpha}+y}}{(\beta+1)^{\alpha+y}} \\
&= \frac{1}{y!} \cdot \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)} \cdot \frac{\beta^y}{(\beta+1)^{\alpha+y}}
\end{aligned}
$$

Though this is a perfectly valid form for the unconditional p.m.f. of Y, we are asked to identify this distribution by name. To do so, we'll simplify our result a bit, by leveraging the fact that both α and y are assumed to be integers:

$$
\begin{aligned}
p_Y(y) &= \frac{1}{y!} \cdot \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)} \cdot \frac{\beta^y}{(\beta+1)^{\alpha+y}} \\
&= \frac{(\alpha+y-1)!}{y! \cdot (\alpha-1)!} \cdot \left(\frac{1}{\beta+1}\right)^\alpha \cdot \left(\frac{\beta}{\beta+1}\right)^y \\
&= \binom{\alpha+y-1}{y} \cdot \left(\frac{1}{\beta+1}\right)^\alpha \cdot \left(1 - \frac{1}{\beta+1}\right)^y
\end{aligned}
$$

which is valid for $y \in \{0, 1, 2, \cdots\}$. Staring at this, we see this is the p.m.f. of a Negative Binomial distribution on $\{0, 1, 2, \cdots\}$; specifically,

$$Y \sim \text{NegBin}\left(\alpha\,,\ \frac{1}{\beta+1}\right)$$

....

Here is an alternate way you can evaluate the integral for this part, using $u-$substitutions. We start with

$$\begin{aligned} p_Y(y) &= \frac{1}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \int_0^\infty \lambda^{(\alpha+y)-1} e^{-\lambda\left(1+\frac{1}{\beta}\right)}\,\mathrm{d}\lambda \\ &= \frac{1}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \int_0^\infty \lambda^{(\alpha+y)-1} e^{-\lambda\left(\frac{\beta+1}{\beta}\right)}\,\mathrm{d}\lambda \end{aligned}$$

The integral closely resembles the definition of the Gamma function:

$$\Gamma(r) := \int_0^\infty t^{r-1} e^{-t}\,\mathrm{d}t$$

However, it is not exactly a Gamma integral since the exponent is not just the variable of integration - it involves a *constant* times the variable of integration. But that's not a problem - we can simply make a $u-$substitution and define the new variable of integration to be precisely whatever is in the exponent! That is:

$$u = \lambda\left(\frac{\beta}{\beta+1}\right) \implies \mathrm{d}u = \frac{\beta}{\beta+1}\,\mathrm{d}\lambda \implies \mathrm{d}\lambda = \frac{\beta+1}{\beta}\,\mathrm{d}u$$

Additionally,

$$\lambda = \left(\frac{\beta}{\beta+1}\right) u$$

and so, substituting into our integral, we have:

$$\begin{aligned} p_Y(y) &= \frac{1}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \int_0^\infty \lambda^{(\alpha+y)-1} e^{-\lambda\left(\frac{\beta+1}{\beta}\right)}\,\mathrm{d}\lambda \\ &= \frac{1}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \int_0^\infty \left[\left(\frac{\beta}{\beta+1}\right) u\right]^{(\alpha+y)-1} e^{-u} \cdot \frac{\beta+1}{\beta}\,\mathrm{d}u \\ &= \frac{1}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \left(\frac{\beta}{\beta+1}\right)^{\alpha+y} \cdot \int_0^\infty u^{(\alpha+y)-1} \cdot e^{-\lambda u}\,\mathrm{d}u \end{aligned}$$

The integral is now exactly the definition of $\Gamma(\alpha+y)$, meaning

$$\begin{aligned} p_Y(y) &= \frac{1}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \left(\frac{\beta}{\beta+1}\right)^{\alpha+y} \cdot \int_0^\infty u^{(\alpha+y)-1} \cdot e^{-\lambda u}\,\mathrm{d}u \\ &= \frac{1}{y!} \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \left(\frac{\beta}{\beta+1}\right)^{\alpha+y} \cdot \Gamma(\alpha+y) \end{aligned}$$

which is precisely what we obtained above, using the other method.

PLEASE NOTE: As was stated in a couple of Canvas Assignments, you were not asked to submit Problem 4 with HW01. However, I am providing solutions to this problem as its material (specifically, conditional expectations and the Law of Iterated Expectations) is fair game for Quiz01. Part (b) [which utilizes the Law of Total Variance] is out-of-scope for Quiz01.

4. Let N be a random variable whose support consists only of natural numbers, and let $\{Y_i\}_{i \geq 0}$ denote a sequence of identically distributed (but not necessarily independent) random variables with mean μ and variance σ^2. Furthermore, define

$$S_N := \sum_{i=1}^{N} Y_i$$

Notice, crucially, that the sum on the RHS above contains a *random* number of terms.

(a) Prove **Wald's Theorem**, which states that $\mathbb{E}[S_N] = \mathbb{E}[N] \cdot \mu$.

Solution: The trick is to appeal to the Law of Iterated Expectations. First:

$$\mathbb{E}[S_N \mid N = n] := \mathbb{E}\left[\sum_{i=1}^{N} Y_i \mid N = n\right] = \mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = n\mu$$

Hence, by definition, $\mathbb{E}[S_N \mid N] = N\mu$ and, by the Law of Iterated Expectations,

$$\mathbb{E}[S_N] = \mathbb{E}[\mathbb{E}[S_N \mid N]] = \mathbb{E}[N\mu] = \mathbb{E}[N] \cdot \mu$$

which is precisely the desired result.

(b) Derive a formula for $\text{Var}(S_N)$. If you need to make an additional assumption, clearly state which assumption(s) need to be made.

Solution: Here, we need to appeal to the Law of Total Variance:

$$\text{Var}(S_N) = \mathbb{E}[\text{Var}(S_N \mid N)] + \text{Var}(\mathbb{E}[S_N \mid N])$$

We know that $\mathbb{E}[S_N \mid N] = N\mu$ (as was derived in the previous part). To compute $\text{Var}(S_N \mid N)$, we write

$$\text{Var}(S_N \mid N = n) = \text{Var}\left(\sum_{i=1}^{N} Y_i \mid N = n\right) = \text{Var}\left(\sum_{i=1}^{n} Y_i\right)$$

at this point, we would like to pull the variance through the sum, However, we can only do so if we assume independence; hence, **let's assume independence**. Then

$$\text{Var}(S_N \mid N = n) = \text{Var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \text{Var}(Y_i) = n\sigma^2 \implies \text{Var}(S_N \mid N) = N\sigma^2$$

Hence, putting everything together:

$$\text{Var}(S_N) = \mathbb{E}[\text{Var}(S_N \mid N)] + \text{Var}(\mathbb{E}[S_N \mid N])$$

$$
\begin{aligned}
&= \mathbb{E}\left[N\sigma^2\right] + \mathsf{Var}\left(N\mu\right) \\
&= \sigma^2 \cdot \mathbb{E}[N] + \mu^2 \mathsf{Var}(N)
\end{aligned}
$$

Again, this is only valid if we assume the Y_i to be i.i.d.; otherwise, this formula does not hold.