

Name: _____

ID:

18440: Probability and Random variables
Quiz 2
Solutions

- You will have 50 minutes to complete this test.
- No calculators, notes, or books are permitted.
- If a question calls for a numerical answer, you do not need to multiply everything out. (For example, it is fine to write something like $(0.9)^7 / (3!2!)$ as your answer.)
- Don't forget to write your name on the top of every page.
- Please show your work and explain your answer. We will not award full credit for the correct numerical answer without proper explanation. Good luck!

Name: _____

ID: _____

Problem 1 (35 points) 3 types of particles (first, second and third type) are created according to $N_1(t), \dots, N_3(t), t \geq 0$ independent Poisson processes with $\lambda = 1$.

- (5 points) What is the law of $\sum_{i=1}^3 N_i(1)$? (give the name and density)
- (5 points) What is the probability that $N_1(1) + N_2(2)$ equal 10?
- (5 points) Write down the density function for the amount of time until a first particle (of any kind) is created.
- (10 points) Compute the probability that after one hour 3 particles are created knowing that no particles of the first kind were.
- (10 points) What is the conditional probability distribution of $(N_1(1) + N_2(2))^4$, knowing $N_1(1) + N_2(2) = 5N_3(1)$?

Proof. (1) The sum of independent Poisson random variables is Poisson with parameter equal to the sum of parameters. Hence $\sum_{i=1}^3 N_i(1) \sim \text{Poisson}(3)$, and so

$$P\left(\sum_{i=1}^3 N_i(1) = k\right) = \frac{e^{-3} 3^k}{k!}.$$

(2) As above we have that $N_1(1) + N_2(2) \sim \text{Poisson}(1 + 2)$, hence

$$P(N_1(1) + N_2(2) = 10) = \frac{e^{-3} 3^{10}}{10!}.$$

(3) Let T_1 be the first arrival time of a particle from $N_1(t) + N_2(t) + N_3(t)$. We know that $N_1(t) + N_2(t) + N_3(t)$ is a Poisson process with rate the sum of the rates, in this case 3. Also the first arrival time of a Poisson process with rate λ is distributed according to $\text{expon}(\lambda)$. So we conclude that

$$f_{T_1}(t) = \begin{cases} 3e^{-3t} & \text{if } t > 0 \\ 0 & \text{otherwise.} \end{cases}$$

(4) The question asks to find $P(N_1(1) + N_2(1) + N_3(1) = 3 | N_1(1) = 0)$. Using Bayes rule and independence we get

$$\begin{aligned} P(N_1(1) + N_2(1) + N_3(1) = 3 | N_1(1) = 0) &= \frac{P(N_1(1) + N_2(1) + N_3(1) = 3, N_1(1) = 0)}{P(N_1(1) = 0)} = \\ &= \frac{P(N_2(1) + N_3(1) = 3, N_1(1) = 0)}{P(N_1(1) = 0)} = \frac{P(N_2(1) + N_3(1) = 3)P(N_1(1) = 0)}{P(N_1(1) = 0)} = \\ &= P(N_2(1) + N_3(1) = 3) = e^{-2} \frac{2^3}{3!}. \end{aligned}$$

(5) We have that $N_1(1)$ and $N_2(2)$ are non-negative integer random variables, so $(N_1(1) + N_2(2))^4$ is a random variable supported on $\{k^4 | k \in \mathbb{Z}_{\geq 0}\}$. So to determine the distribution it suffices to find $P((N_1(1) + N_2(2))^4 = k^4 | N_1(1) + N_2(2) = 5N_3(1)) = P(N_1(1) + N_2(2) = k | N_1(1) + N_2(2) = 5N_3(1))$. Using Bayes rule and independence we get

$$\begin{aligned} P(N_1(1) + N_2(2) = k | N_1(1) + N_2(2) = 5N_3(1)) &= \frac{P(N_1(1) + N_2(2) = k, N_1(1) + N_2(2) = 5N_3(1))}{P(N_1(1) + N_2(2) = 5N_3(1))} = \\ &= \frac{P(N_1(1) + N_2(2) = k, k = 5N_3(1))}{P(N_1(1) + N_2(2) = 5N_3(1))} = \frac{P(N_1(1) + N_2(2) = k)P(5N_3(1) = k)}{P(N_1(1) + N_2(2) = 5N_3(1))}. \end{aligned}$$

Now we estimate the denominator and get

$$B = P(N_1(1) + N_2(2) = 5N_3(1)) = \sum_{m=0}^{\infty} P(N_1(1) + N_2(2) = 5m, N_3(1) = m) =$$

$$\sum_{m=0}^{\infty} P(N_1(1) + N_2(2) = 5m)P(N_3(1) = m) = \sum_{m=0}^{\infty} e^{-3} \frac{3^{5m}}{(5m)!} e^{-1} \frac{1^m}{m!}$$

On the other hand we have for the numerator

$$A(k) = P(N_1(1) + N_2(2) = k)P(5N_3(1) = k) = \begin{cases} e^{-3} \frac{3^{5m}}{(5m)!} e^{-1} \frac{1^m}{m!} & \text{if } k = 5m \\ 0 & \text{if } 5 \nmid k \end{cases} .$$

We thus conclude that the conditional distribution is

$$P((N_1(1) + N_2(2))^4 = k^4 | N_1(1) + N_2(2) = 5N_3(1)) = \frac{A(k)}{B},$$

and 0 otherwise.

□

Name: _____

ID: _____

Problem 2 (20 points) Let X_1, X_2, X_3, X_4 be independent uniform variables on $[0, 1]$. Write $Y = X_1 + X_2 + X_3$ and $Z = X_1 + X_2$.

- (5 points) Compute the density function of Z .
- (5 points) Compute the probability that $\{Y < 1/2\}$.
- (5 points) Compute $Cov(Z, Y)$.
- (5 points) Are Z and Y independent? Why?

Proof. (1) Since X_i are supported in $[0, 1]$ we know $X_1 + X_2$ is supported in $[0, 2]$. Next we have that

$$f_Z(y) = \begin{cases} \int_{y-1}^1 f_{X_1}(a)f_{X_2}(y-a)da = 2-y & \text{if } 2 \geq y \geq 1 \\ \int_0^y f_{X_1}(a)f_{X_2}(y-a)da = y & \text{if } y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(2) We have

$$\begin{aligned} P(Y < 1/2) &= P(X_1 + X_2 + X_3 < 1/2) = \int_0^{1/2} \int_0^{1/2-x_1} \int_0^{1/2-x_1-x_2} dx_3 dx_2 dx_1 = \\ &= \int_0^{1/2} \int_0^{1/2-x_1} (1/2 - x_1 - x_2) dx_2 dx_1 = \\ &= \int_0^{1/2} \frac{1}{2} (1/2 - x_1)^2 dx_1 = \frac{1}{2} \int_0^{1/2} u^2 du = \frac{1}{48}. \end{aligned}$$

(3) By linearity of expectation we know

$$E[Z] = E[X_1] + E[X_2] = 1/2 + 1/2 = 1 \text{ and similarly } E[Y] = 3/2.$$

Next we have

$$\begin{aligned} E[Z Y] &= \int_0^1 \int_0^1 \int_0^1 (x_1 + x_2)(x_1 + x_2 + x_3) dx_3 dx_2 dx_1 = \int_0^1 \int_0^1 (x_1 + x_2)^2 + \frac{1}{2}(x_1 + x_2) dx_2 dx_1 = \\ &= \int_0^1 x_1^2 + \frac{3}{2}x_1 + \frac{7}{12} = \frac{1}{3} + \frac{3}{4} + \frac{7}{12} = \frac{20}{12}. \end{aligned}$$

Consequently,

$$Cov(Z, Y) = E[Z Y] - E[Z]E[Y] = \frac{5}{3} - \frac{3}{2} = \frac{1}{6}.$$

(4) It is clear that Y and Z are not independent, as otherwise their covariance would be 0, which we saw to not be the case. Intuitively, we have $Y > Z$ so knowing Z we have a lower bound on Y so the two cannot be independent.

□

Name: _____

ID: _____

Problem 3 (25 points) Consider a sequence of independent tosses of a coin that is biased so that it comes up heads with probability $2/3$ and tails with probability $1/3$. Let X_i be one if the i -th toss is head, and 0 otherwise.

- (1) (5 points) Compute $\mathbb{E}[X_i]$ and $Var(X_i)$.
- (2) (5 points) Compute $Var(X_1 + 2X_2 + 3X_3 + 4X_4 + 5X_5)$
- (3) (5 points) Let $Y = \sum_{i=1}^{40000} X_i$. Approximate the probability that $Y \geq 27667$.
- (4) (5 points) Let $Z = \sum_{i=1}^{20000} X_i - \sum_{i=20001}^{40000} X_i$. What can you say about the distribution of $Z/200$?
- (5) (5 points) Show that $(Y/200, Z/200)$ are approximately independent.

Proof. (1) Set $p = 2/3$. Then

$$E[X_i] = 1 \times P(\text{heads on } i\text{th toss}) + 0 = p,$$

$$Var(X_i) = E[X_i^2] - E[X_i]^2 = E[X_i] - E[X_i]^2 = p(1 - p).$$

(2) We use that for independent random variables X, Y one has

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y),$$

to get in view of part (1)

$$Var(X_1 + 2X_2 + 3X_3 + 4X_4 + 5X_5) = \sum_{i=1}^5 i^2 Var(X_i) = p(1 - p)(1 + 4 + 9 + 16 + 25) = 55p(1 - p).$$

(3) We have using part (1) and linearity of expectation and variance for independent variables

$$E[Y] = 40000 \times p = \frac{80000}{3} \approx 26667 \text{ as well as } Var(Y) = 40000 \times p(1 - p) = \frac{80000}{9} \approx 8889.$$

Consequently we have by De Moivre - Laplace

$$P\left(\frac{Y - 26667}{\sqrt{8889}} \geq a\right) \approx 1 - \Phi(a),$$

where Φ denotes the Gaussian cdf, i.e.

$$\Phi(y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Consequently, we have

$$P(Y \geq 27667) = P\left(\frac{Y - 26667}{\sqrt{8889}} \geq \frac{27667 - 26667}{\sqrt{8889}}\right) \approx 1 - \Phi(a),$$

where $a = \frac{1000}{\sqrt{8889}}$.

(4) Let $T_1 = \sum_{i=1}^{20000} X_i$ and $T_2 = \sum_{i=20001}^{40000} X_i$. Similarly to our work above we know that $E[T_i] = 40000/3 = a$ and $Var(T_i) = 40000/9 = b$. By De Moivre Laplace we know that the distribution of

$$\frac{T_i - 40000/3}{200/3},$$

is approximately $Gaussian(0, 1) = N(0, 1)$. Consequently the distribution of $T_i/200$ is approximately $N(200/3, 1/9)$. Also it is clear T_i are independent as they depend on disjoint collection of tosses. Thus

$$Z/200 = T_1 - T_2 \approx A - B,$$

where A, B are independent $N(200/3, 1/9)$. Using that if $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, one has

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2),$$

we conclude that

$$A - B \sim N(0, 2/9).$$

So the distribution of $Z/200$ is approximately $N(0, 2/9)$.

(5) We have $Y/200 = T_1/200 + T_2/200$ so we have that the distribution of $Y/200$ is approximately $N(400/3, 2/9)$. We know that two Gaussian random variables are independent if and only if the covariance is 0. Since $Z/200$ and $Y/200$ have approximately Gaussian distributions they would be approximately independent if their covariance is 0. But we have

$$\begin{aligned} \text{Cov}(Y/200, Z/200) &= \frac{1}{40000}(E[YZ] - E[Y]E[Z]) = \frac{1}{40000}(E[(T_1+T_2)(T_1-T_2)] - E[T_1+T_2]E[T_1-T_2]) = \\ &= \frac{1}{40000}(E[T_1^2] - E[T_2^2] - E[T_1]^2 + E[T_2]^2) = 0, \end{aligned}$$

where in the last line we used that T_1 and T_2 have the same distribution and hence the same moments.

□

Name: _____

ID: _____

Problem 4 (10 points) Let X and Y be continuous random variables with joint probability density

$$f_{X,Y}(a,b) = \begin{cases} 10a^3 + 21b^5 & \text{on } a+b \leq 1, a \geq 1, b \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(1) (5 points) Find the probability distribution of $Z = 2Y + X$.

(2) (5 points) Find the probability distribution of $W = (X + Y)^2 + 1$.

Proof. (1): We calculate the density of Z . Since X, Y are supported on $[0, 1]$ we know Z is supported on $[0, 3]$. Then we have

$$f_Z(x) = \int_0^\infty f_{X,Y}(x-2b, b) db = \int_{\max(0, (x-1)/2)}^{\min(1, x/2)} 10(x-2b)^3 + 21b^5 db.$$

Setting $B = \max(0, (x-1)/2)$ and $A = \min(1, x/2)$, above becomes

$$\frac{7A^6 - 7B^6}{2} + (-20A^4 + 20B^4) + 10x^3(A - B) + (-60)x^2(A^2/2 - B^2/2) + 40x(A^3 - B^3).$$

(2): We calculate the density of W . Again we observe W is supported on $[1, 2]$. Then we have for $x \geq 1$

$$f_W(x) = \int_0^\infty f_{X,Y}(a, \sqrt{x-1} - a) da,$$

otherwise f_W is 0. We calculate the above to be

$$\int_{\max(0, \sqrt{x-1}-1)}^{\min(1, \sqrt{x-1})} 10a^3 + 21(\sqrt{x-1} - a)^5 da.$$

Setting $A = \min(1, \sqrt{x-1})$ and $B = \max(0, \sqrt{x-1} - 1)$ we get the above equals

$$\frac{10}{4}(A^4 - B^4) + \sum_{i=0}^5 \binom{5}{i} \frac{(-1)^i 21(\sqrt{x-1})^{5-i}}{i+1} (A^{i+1} - B^{i+1}).$$

□

Name: _____

ID: _____

Problem 5 (10 points) A model for the movement of a stock supposes that if the present price of the stock is s , then after one period, it will be either us with probability p , or ds with probability $1 - p$. Assuming that successive movements are independent, approximate the probability that the stock's price will be up at least 30 percent after the next 1000 periods if $u = 1.012$, $d = 0.990$, $p = .52$.

Proof. Let X be the number of up jumps of the stock price and $1000 - X$ the number of down jumps. X is Binomial(1000,p). Then we have that the price of the stock after 1000 periods is

$$S_{final} = u^X d^{1000-X} s.$$

Consequently, we are interested in

$$P(S_{final} \geq 1.3s) = P(u^X d^{1000-X} \geq 1.3),$$

which upon taking logs is just

$$P(X \log(u) + (1000 - X) \log(d) \geq \log(1.3)).$$

We fix

$$\log(u) = a \quad \log(d) = b.$$

Then our probability becomes

$$P(Xa + (1000 - X)b \geq \log(1.3)) = P(X \geq \frac{\log(1.3) - 1000b}{a - b}).$$

Using De Moivre - Laplace formula we know that

$$P\left(\frac{X - 1000p}{\sqrt{1000p(1-p)}} \geq y\right) \approx 1 - \Phi(y),$$

where

$$\Phi(y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Consequently we have that if $C = \frac{\log(1.3) - 1000b}{a - b}$ then

$$P(X \geq C) = P\left(\frac{X - 1000p}{\sqrt{1000p(1-p)}} \geq \frac{C - 1000p}{\sqrt{1000p(1-p)}}\right) \approx \Phi(C'),$$

where

$$C' = \frac{C - 1000p}{\sqrt{1000p(1-p)}}.$$

This suffices for the proof. □