Probability and Statistics | MATH-235
Autumn 2025 | Mats J. Stensrud

Problem Sheet 10 [1]

Based on Chapters 7.3, 9.1-9.4, 9.5, and 10.1 of the course book.

Optional Revision Problems

Exercise 1. Let X and Y have joint PDF $f_{X,Y}(x, y) = x + y$, for $0 < x < 1$ and $0 < y < 1$.

1. Check that this is a valid joint PDF.
2. Find the marginal PDFs of X and Y.
3. Are X and Y independent?
4. Find the conditional PDF of Y given $X = x$.

Solution 1.

1. The PDF is valid if it is non-negative, and integrates to 1 over the support. Since both $x, y > 0$, their sum is non negative, and
$$\int_0^1 \int_0^1 x + y \, \mathrm{d}x\mathrm{d}y = \int_0^1 \frac{1}{2} + y \, \mathrm{d}y = \frac{1}{2} + \frac{1}{2} = 1,$$
so $f_{X,Y}(x, y) = x + y$ is a valid PDF.
2. By definition $f_X(x) = \int_0^1 f_{X,Y}(x, y)\mathrm{d}y = \int_0^1 x + y \, \mathrm{d}y = x + \frac{1}{2}$, for $0 < x < 1$, and 0 otherwise. Similarly $f_Y(y) = \int_0^1 f_{X,Y}(x, y)\mathrm{d}x = y + \frac{1}{2}$, for $0 < y < 1$, and 0 otherwise.
3. Since $f_{X,Y}(x, y) = x + y \neq (\frac{1}{2} + x)(\frac{1}{2} + y) = f_X(x)f_Y(y)$, X and Y are not independent.
4. By definition $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{x+y}{x+\frac{1}{2}}$ for $0 < y < 1$ and $0 < x < 1$ and it is 0 otherwise.

1 Week 10 Exercises

Exercise 2.

1. Let X and Y be Bernoulli r.v.s, possibly with different parameters. Show that if X and Y are uncorrelated, then they are independent.
2. Give an example of three Bernoulli r.v.s such that each pair of them is uncorrelated, yet the three r.v.s are dependent.

[1] Exercises are based on the coursebook Statistics 110: Probability by Joe Blitzstein

Solution 2. 1. If X and Y are uncorrelated, that implies that their covariance is zero:

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0 \implies E(XY) = E(X)(Y).$$

The expectation of Bernoulli random variables are just their parameters, that is, if $X \sim Bern(p)$ and $Y \sim Bern(q)$, then $E(X) = p$ and $E(Y) = q$. These parameters stand for the probability of success, so $p = P(X = 1)$ and $q = P(Y = 1)$. Similarly,

$$\begin{aligned} E(XY) &= 1 \cdot 1 \cdot P(X = 1, Y = 1) + 0 \cdot 1 \cdot P(X = 0, Y = 1) \\ &\quad + 1 \cdot 0 \cdot P(X = 1, Y = 0) + 0 \cdot 0 \cdot P(X = 0, Y = 0) \\ &= P(X = 1, Y = 1), \end{aligned}$$

therefore, if they are uncorrelated, $P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = p \cdot q$. To show independence, we have to show that the joint probability can be factorized to the marginals, for all other combinations of the values.

By the law of total probability $P(X = x) = P(X = x, Y = 0) + P(X = x, Y = 1)$, so

$$P(X = 1, Y = 0) = P(X = 1) - P(X = 1, Y = 1) = p - pq = p(1 - q) = P(X = 1)P(Y = 0).$$

Similarly,

$$P(X = 0, Y = 1) = P(Y = 1) - P(X = 1, Y = 1) = q - pq = q(1 - p) = P(Y = 1)P(X = 0),$$

and finally

$$\begin{aligned} P(X = 0, Y = 0) &= P(X = 0) - P(X = 0, Y = 1) = (1 - p) - q(1 - p) = (1 - p)(1 - q) \\ &= P(X = 0)P(Y = 0). \end{aligned}$$

For any $x, y \notin \{0, 1\}$, $P(X = x, Y = y) = 0 = P(X = x)P(Y = y)$, so the joint probability can be written as the product of marginals, therefore X and Y are independent.

2. Let X and Y be two independent Bernoulli random variables, with parameter $\frac{1}{2}$. Define Z as

$$Z = \begin{cases} 1 & \text{if } X \neq Y \\ 0 & \text{if } X = Y. \end{cases}$$

This implies that the marginal distribution of Z is $Bern(\frac{1}{2})$ as well, because $P(X = Y) = \frac{1}{2}$, that you can show by the axioms of probability. Then the joint distribution of X, Y, and Z have the following probabilities:

$$\Pr(X = 0, Y = 0, Z = 0) = \frac{1}{4}, \quad \Pr(X = 1, Y = 1, Z = 0) = \frac{1}{4},$$

$$\Pr(X = 1, Y = 0, Z = 1) = \frac{1}{4}, \quad \Pr(X = 0, Y = 1, Z = 1) = \frac{1}{4},$$

and the joint probability is zero for any other combination of these random variables. The three random variables are dependent, as $P(X = 0, Y = 0, Z = 0) = \frac{1}{4} \neq \frac{1}{8} = P(X = 0)P(Y = 0)P(Z = 0)$, or you can argue that knowing the value of X and Y determines the value of Z

By definition X and Y are independent. Moreover both pairs of X and Z and Y and Z are independent, as $P(Z = z|Y = y) = P(Z = z) = \frac{1}{2}$, because the value of Z then will be determined by the value of X, that is $Bern(\frac{1}{2})$.

You can also show the independence of the pairs by using the result from part 1., as follows:

The expectation of each random variable is $\frac{1}{2}$, as they are $Bern(\frac{1}{2})$.

For the expectation of each pair, first, we need the PMF for the pairs. For example for X and Z, that is

$$
\begin{aligned}
P(X = 0, Z = 0) &= P(X = 0, Y = 0, Z = 0) + P(X = 0, Y = 1, Z = 0) = \frac{1}{4} + 0 = \frac{1}{4} \\
P(X = 1, Z = 0) &= P(X = 1, Y = 0, Z = 0) + P(X = 1, Y = 1, Z = 0) = 0 + \frac{1}{4} = \frac{1}{4} \\
P(X = 0, Z = 1) &= P(X = 0, Y = 0, Z = 1) + P(X = 0, Y = 1, Z = 1) = 0 + \frac{1}{4} = \frac{1}{4} \\
P(X = 1, Z = 1) &= P(X = 1, Y = 0, Z = 1) + P(X = 1, Y = 1, Z = 1) = \frac{1}{4} + 0 = \frac{1}{4}.
\end{aligned}
$$

Then the expectation is

$$E(XZ) = \frac{1}{4}(0 \cdot 0 + 1 \cdot 0 + 0 \cdot 1 + 1 \cdot 1) = \frac{1}{4}.$$

Therefore $E(XZ) = E(X)E(Z)$, so they are uncorrelated. By part 1. this implies independence. The correlation being equal to zero, and then independence, follows similarly for the other two pairs.

The main takeaway from this exercise is that *pairwise independence* **does not imply** *independence in general*.

Exercise 3. Show that $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$

Solution 3. Using the expression $Cov(X, Y) = E(XY) - E(X)E(Y)$ for the covariance, we have

$$
\begin{aligned}
Cov(X + Y, Z) &= E((X + Y)Z) - E(X + Y)E(Z) \\
&= E(XZ + YZ) - E(X + Y)E(Z) \\
&= E(XZ) + E(YZ) - (E(X) + E(Y))E(Z) \\
&= (E(XZ) - E(X)E(Z)) + (E(YZ) - E(Y)E(Z)) \\
&= \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z).
\end{aligned}
$$

To get the second line we just opened the brackets, in the third line we used the linearity of the expectation, in the fourth we just reordered the terms, and finally, we used again the expression above for the covariance.

Exercise 4. Show that for any two random variables X and Y,

$$-1 \leq \mathrm{Corr}(X, Y) \leq 1.$$

Hint: Consider the variance of $X + Y$ and $X - Y$, respectively.

Solution 4. Assume that X and Y have variance 1, as the correlation is scale independent (if in doubt, write out the correlation of X and Y, and the correlation of cX and Y for some $c \in \mathbb{R}$, using the definition of the correlation). In this case $\text{Cov}(X,Y) = \text{Corr}(X,Y)$, so we just want to show that $\rho = \text{Cov}(X,Y) \in [-1,1]$. Then,

$$Var(X+Y) = Var(X) + Var(Y) + 2\text{Cov}(X,Y) = 2 + 2\rho \geq 0,$$
$$Var(X-Y) = Var(X) + Var(Y) - 2\text{Cov}(X,Y) = 2 - 2\rho \geq 0,$$

since variances are always non-negative. From the first line, it follows that $-1 \leq \rho$, and from the second that $\rho \leq 1$, as desired.

Exercise 5. Optional Exercise You get to choose between two envelopes, each of which contains a check for some positive amount of money. Unlike in the two-envelope paradox, it is not given that one envelope contains twice as much money as the other envelope. Instead, assume that the two values were generated independently from some distribution on the positive real numbers, with no information given about what that distribution is.

After picking an envelope, you can open it and see how much money is inside (call this value x), and then you have the option of switching. As no information has been given about the distribution, it may seem impossible to have better than a 50% chance of picking the better envelope. Intuitively, we may want to switch if x is "small" and not switch if x is "large", but how do we define "small" and "large" in the grand scheme of all possible distributions? [The last sentence was a rhetorical question.]

Consider the following strategy for deciding whether to switch. Generate a threshold $T \sim \textit{Expo}(1)$, and switch envelopes if and only if the observed value x is less than the observed value of T. Show that this strategy succeeds in picking the envelope with more money with probability strictly greater than 1/2.

Hint: Let t be the value of T (generated by a random draw from the $\textit{Expo}(1)$ distribution). First explain why the strategy works very well if t happens to be in between the two envelope values, and does no harm in any case (i.e., there is no case in which the strategy succeeds with probability strictly less than 1/2).

Solution 5. Let a be the smaller value of the two envelopes and b be the larger value (assume $a < b$ since in the case $a = b$ it makes no difference which envelope is chosen!). Let G be the event that the strategy succeeds and A be the event that we pick the envelope with a initially. Then

$$P(G|A) = P(T > a) = 1 - (1 - e^{-a}) = e^{-a},$$

and

$$P(G|A^c) = P(T \leq b) = 1 - e^{-b}.$$

Thus, the probability that the strategy succeeds is

$$\frac{1}{2}e^{-a} + \frac{1}{2}(1 - e^{-b}) = \frac{1}{2} + \frac{1}{2}(e^{-a} - e^{-b}) > \frac{1}{2},$$

because $e^{-a} - e^{-b} > 0$.

Exercise 6. Let $X \sim \textit{Expo}(\lambda)$. Find $E(X|X < 1)$ in two different ways:

1. By calculus, working with the conditional PDF of X given $X < 1$.

   Hint: For the conditional PDF use the formula under Definition 9.1.1 in the book.

2. Without calculus, by expanding $E(X)$ using the law of total expectation.

   Hint: Use the memoryless property of the exponential, in particular that if X is Exponential, then $E(X|X \geq a) = E(X) + a$ (if you don't see this, look at Example 9.1.8, that describes this property for the discrete case).

Solution 6. 1. Since $X < 1$ is an event, the conditional PDF can be written as

$$f_{X|X<1}(x|X<1) = \frac{P(X<1|X=x)f_X(x)}{P(X<1)}.$$

Since $P(X < 1|X = x)$ is 0 if $x > 1$ and 1 otherwise, and the support of the exponential distribution is $[0, \infty)$, the conditional PDF will be non-zero only on $[0, 1]$. The last thing is to find $P(X < 1)$, but that is by definition $F_X(1) = 1 - e^{-\lambda \cdot 1} = 1 - e^{-\lambda}$.

So the conditional PDF is

$$f_{X|X<1}(x|X<1) = \frac{\lambda e^{-\lambda x}}{1-e^{-\lambda}},$$

for $0 < x < 1$ and 0 otherwise.

The conditional expectation by definition is

$$E(X|X<1) = \int_0^1 x \frac{\lambda e^{-\lambda x}}{1-e^{-\lambda}}\, \mathrm{d}x = \frac{1}{1-e^{-\lambda}} \int_0^1 \lambda e^{-\lambda x} x\, \mathrm{d}x = \frac{1}{1-e^{-\lambda}} \left(\left[-e^{-\lambda x} x\right]_0^1 + \int_0^1 e^{-\lambda x}\, \mathrm{d}x \right)$$
$$= \frac{1}{1-e^{-\lambda}} \left(-e^{-\lambda} + \left[-\frac{e^{-\lambda x}}{\lambda} \right]_0^1 \right) = \frac{1}{1-e^{-\lambda}} \left(-e^{-\lambda} + \frac{1}{\lambda}(1-e^{-\lambda}) \right) = \frac{1}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}}.$$

You can check that this function is decreasing in λ and for some $\lambda_n \to 0^+$ it is equal to 0.5, which means that $E(X|X < 1) < 1$, that is a good sanity check, as we condition on $X < 1$. For example, imagine that you conduct an experiment and you discard all the results that are greater than 1. Then from all the remaining experimental results, you calculate the average. In this scenario, it would be impossible to get an average that is greater than 1.

2. By the law of total expectation

$$E(X) = E(X|X \geq 1)P(X \geq 1) + E(X|X<1)P(X<1).$$

Using the hint, and that $P(X \geq 1) = 1 - P(X < 1) = e^{-\lambda}$, and that the expectation of an $Expo(\lambda)$ is $1/\lambda$, we can rewrite the above equation as

$$\frac{1}{\lambda} = (\frac{1}{\lambda} + 1)e^{-\lambda} + E(X|X<1)(1-e^{-\lambda}),$$

so after rearranging

$$E(X|X<1) = \frac{\frac{1}{\lambda}(1-e^{-\lambda}) - e^{-\lambda}}{1-e^{-\lambda}} = \frac{1}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}},$$

just as previously shown, but without any integrals involved.

Exercise 7. A fair 6-sided die is rolled once. Find the expected number of additional rolls needed to obtain a value at least as large as that of the first roll.

Hint: If you get a conditional expectation somewhere in your solution, try to think about what its distribution could be, instead of "brute-force" calculating the expectation.

Solution 7. Denote the number of additional rolls with L, and the value of the first roll with X. By the law of total expectation

$$E(L) = E(L|X = 1)P(X = 1) + \cdots + E(L|X = 6)P(X = 6).$$

Since the die is fair $P(X = 1) = \cdots = P(X = 6) = 1/6$. For the conditional expectations, we just have to realize that $L|X = a$ has First Success distribution with parameter $p = \frac{6-a+1}{6}$, as we have a sequence of independent Bernoulli trials (rolls) and we are asking when do we have the first success. The success probability is determined by the first roll, as if we roll 1 first ($X = 1$), any roll will be a success, so $p = 1$, but if we roll a 5 first ($X = 5$), then only 5 and 6 are a success, so we have a success probability of $\frac{2}{6} = \frac{1}{3}$. The expectation of a First Success distribution with parameter p is $\frac{1}{p}$, therefore

$$E(L) = \frac{1}{6}\left(\frac{6}{6} + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1}\right) = \frac{147}{60} = 2.45,$$

is the expected number of additional rolls.

Exercise 8. Show that $E((Y - E(Y|X))^2|X) = E(Y^2|X) - (E(Y|X))^2$, so these two expressions for $Var(Y|X)$ agree.

Hint for the variance: Adding a constant (or something acting as a constant) does not affect variance.

Solution 8. Define $E(Y|X) = g(X)$. Then we have

$$E((Y - E(Y|X))^2|X) = E(Y^2 - 2Yg(X) + g(X)^2|X) = E(Y^2|X) - 2E(Yg(X)|X) + E(g(X)^2|X),$$

where we used the linearity of the expectation. By taking out what is known $E(Yg(X)|X) = g(X)E(Y|X) = g(X)^2, E(g(X)^2|X) = g(X)^2$, so the right-hand side of the above equation is $E(Y^2|X) - 2g(X)^2 + g(X)^2 = E(Y^2|X) - g(X)^2$. By plugging in the formula from the definition of $g(X)$, we get the desired result.

Exercise 9. Show that if $E(Y|X) = c$ is a constant, then X and Y are uncorrelated. **Hint:** Use Adam's law to find $E(Y)$ and $E(XY)$.

Solution 9. Two random variables are uncorrelated if their covariance is 0, i.e. if $E(XY) - E(X)E(Y) = 0$, then X and Y are uncorrelated.

By Adam's law (law of iterated expectation/tower property)

$$E(Y) = E(E(Y|X)) = E(c) = c,$$

where we used the initial assumption and that the expectation of a constant is the constant itself.

Applying Adam's law (law of iterated expectation/tower property) again

$$E(XY) = E(E(XY|X)) = E(XE(Y|X)) = E(Xc) = cE(X),$$

where for the second equality we used "taking out what's known", then the initial assumption, and finally the linearity of expectation.

Putting everything together

$$E(XY) - E(X)E(Y) = cE(X) - E(X)c = 0,$$

as wanted, so X and Y are uncorrelated.

Exercise 10. Joe will read $N \sim Pois(\lambda)$ books next year. Each book has a $G \sim Pois(\mu)$ number of pages, with book lengths independent of each other and independent of N.

1. Find the expected number of book pages denoted by T, that Joe will read next year.

   Hint: $T \neq N \cdot G$, just consider whether it is possible for Joe to read a prime number of pages in total, even if he read more than one book.

2. ~~Find the variance of the number of book pages Joe will read next year.~~

Solution 10. 1. By Adam's law/tower property, $E(T) = E(E(T|N))$. So we should find $E(T|N)$ first. Since the total number of pages Joe read is just the sum of the pages from the books he read, we can rewrite $T = \sum_{i=1}^{N} G_i$, where $G_i \overset{iid}{\sim} Pois(\mu)$. Using the definition of conditional expectation, given a random variable (Def 9.2.1), if $g(n) := E(T|N = n)$, then $E(T|N) = g(N)$. Conditioning on $N = n$, our expectation simplifies to

$$E(T|N = n) = E\left(\sum_{i=1}^{N} G_i \middle| N = n\right) = E\left(\sum_{i=1}^{n} G_i \middle| N = n\right) = nE(G|N = n),$$

where in the last step we used the linearity of expectation, and that all G_i have the same distribution, hence the same (conditional) mean. As G and N are independent, we can write the right-hand side as $nE(G)$, and because $G \sim Pois(\mu)$, this is equal to $n\mu$. Then, $g(n) = n\mu$, implying that $E(T|N) = g(N) = N\mu$.

Substituting back $E(T) = E(E(T|N)) = E(N\mu) = \mu E(N) = \mu\lambda$, as $N \sim Pois(\lambda)$.

2. *Cliffhanger until next week.*

Had we used **wrongly** the relationship $T = N \cdot G$, we would have gotten the same expectation, but it would be still incorrect to say that. To see that T and $N \cdot G$ do not have the same distribution, calculate the variance of $N \cdot G$, and compare to $Var(T)$.