1. Graphical Density

Figure 1 shows the joint density $f_{X,Y}$ of the random variables $X$ and $Y$.

(a) Find $A$ and sketch $f_X$, $f_Y$, and $f_{X|X+Y\leq3}$.
(b) Find $E[X \mid Y = y]$ for $1 \leq y \leq 3$ and $E[Y \mid X = x]$ for $1 \leq x \leq 4$.
(c) Find $\text{cov}(X,Y)$.

Solution:

(a) The integration over the total shown area should be 1 so $2A + 2A + 2A = 1$ so $A = 1/6$. To spell this out in more detail,

\[
1 = \int_{1}^{3} \int_{1}^{2} A \, dx \, dy + \int_{2}^{3} \int_{2}^{3} 2A \, dx \, dy + \int_{1}^{3} \int_{3}^{4} A \, dx \, dy \\
= 2A + 2A + 2A = 6A.
\]

We find the densities as follows. $X$ is clearly uniform in intervals $(1, 2)$, $(2, 3)$, and $(3, 4)$. The probability of $X$ being in any of these intervals is $2A = 1/3$ so

\[
f_X(x) = \frac{1}{3} \{1 \leq x \leq 4\}.
\]
Y is uniform in intervals (1, 2) and (2, 3). The probability of the first interval is 1/3 and the probability of being in second one is 2/3. So

\[ f_Y(y) = \frac{1}{3} \mathbb{1}_{\{1 \leq y \leq 2\}} + \frac{2}{3} \mathbb{1}_{\{2 < y \leq 3\}}. \]

Finally, given that \( X + Y \leq 3 \), \((X,Y)\) is chosen randomly in the triangle constructed by \((1,1), (1,2), (2,1)\). Thus,

\[ f_{X|X+Y\leq3}(x) = \int_{1}^{3-x} 2 \, dy = 2(2 - x) \mathbb{1}_{\{1 \leq x \leq 2\}}. \]

Sketching the densities is then straightforward.

(b) Given any value of \( y \in [1,3] \), \( X \) has a symmetric distribution with respect to the line \( x = 2.5 \). Thus, \( \mathbb{E}[X \mid Y = y] = 2.5 \) for all \( y, 1 \leq y \leq 3 \). To calculate \( \mathbb{E}[Y \mid X = x] \), we consider two cases:

(a) \( 2 \leq x \leq 3 \), then \( \mathbb{E}[Y \mid X = x] = 2.5 \),

(b) \( 1 \leq x < 2 \) or \( 3 < x \leq 4 \), then \( \mathbb{E}[Y \mid X = x] = 2 \).

(c) Since \( \mathbb{E}[X \mid Y = y] = \mathbb{E}[X] \) we have

\[ \mathbb{E}[XY] = \int_{1}^{3} \mathbb{E}[XY \mid Y = y] f_Y(y) \, dy = \int_{1}^{3} y f_Y(y) \mathbb{E}[X] \, dy = \mathbb{E}[X] \mathbb{E}[Y]. \]

So the covariance is 0.

2. Conditional Distribution of a Poisson Random Variable with Exponentially Distributed Parameter

Let \( X \) have a Poisson distribution with parameter \( \lambda > 0 \). Suppose \( \lambda \) itself is random, having an exponential density with parameter \( \theta > 0 \).

(a) Show that

\[ \mathbb{E}(\lambda^k) = \frac{k!}{\theta^k}, \quad k \in \mathbb{N} \]

(b) What is the distribution of \( X \)?

(c) Determine the conditional density of \( \lambda \) given \( X = k \), where \( k \in \mathbb{N} \).

Solution:

(a) \( \mathbb{E}(\lambda^k) = \int_{0}^{\infty} x^k e^{-\theta x} \, dx \). Integrating by parts, with proper limits,

\[ \mathbb{E}[\lambda^k] = \int_{0}^{\infty} x^k \theta e^{-\theta x} \, dx \]

\[ = -x^k \exp(-\theta x) \bigg|_{x=0}^{\infty} + k \int_{0}^{\infty} x^{k-1} \exp(-\theta x) \, dx \]

\[ = \frac{k}{\theta} \int_{0}^{\infty} x^{k-1} \theta \exp(-\theta x) \, dx, \]
Continuing, and with the base case \( E(\lambda) = \frac{1}{\theta} \), we get
\[
E(\lambda^k) = \frac{k!}{\theta^k}.
\]

(b) The PDF of \( \lambda \) is:
\[
f(\lambda) = \theta \exp(-\theta\lambda) \mathbb{1}_{\{\lambda > 0\}}.
\]

The PMF of \( X \) conditioned on \( \lambda \) is
\[
P(X = k \mid \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k \in \mathbb{N}.
\]

Applying the total law of probability yields, for \( k \in \mathbb{N} \),
\[
P(X = k) = \int_{0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} \theta \exp(-\theta\lambda) \, d\lambda
\]
\[
= \frac{\theta}{(1 + \theta)k!} \int_{0}^{\infty} \lambda^k (1 + \theta) \exp(-(1 + \theta)\lambda) \, d\lambda = \frac{\theta}{(1 + \theta)^{k+1}},
\]

because the last integral is \( E[Y^k] \) when \( Y \sim \text{Exponential}(1 + \theta) \), which is \( k!/(1 + \theta)^k \).

(c)
\[
f(\lambda \mid X = k) = \frac{P(X = k \mid \lambda) f(\lambda)}{P(X = k)} = \frac{e^{-(1+\theta)\lambda} \lambda^k (1 + \theta)^{k+1}}{k!}, \quad \lambda > 0.
\]

To understand the above equation, think about the analogy to Bayes Law. Remember here that \( \theta \) is fixed and \( \lambda \) is the argument. You should check that the integral of this over \([0, \infty)\) is 1.

3. Gaussian Densities

(a) Let \( X_1 \sim \mathcal{N}(0, 1) \), \( X_2 \sim \mathcal{N}(0, 1) \), where \( X_1 \) and \( X_2 \) are independent. Convolve the densities of \( X_1 \) and \( X_2 \) to show that \( X_1 + X_2 \sim \mathcal{N}(0, 2) \).

(b) Let \( X \sim \mathcal{N}(0, \sigma^2) \); find \( E[X^n] \) for \( n \in \mathbb{N} \).

(c) Let \( Z \sim \mathcal{N}(0, 1) \). For a random vector \((X_1, \ldots, X_n)\) where \( n \) is a positive integer and \( X_1, \ldots, X_n \) are real-valued random variables, the expectation of \((X_1, \ldots, X_n)\) is the vector of elementwise expectations of each random variable and the covariation matrix of \((X_1, \ldots, X_n)\) is the \( n \times n \) matrix whose \((i, j)\) entry is \( \text{cov}(X_i, X_j) \) for all \( i, j \in \{1, \ldots, n\} \). Find the mean and covariance matrix of \((Z, \mathbb{1}\{Z > c\})\) in terms of \( \phi \) and \( \Phi \), the standard Gaussian PDF and CDF respectively.

\textbf{Solution:}
(a) **Note:** These solutions are for the more general case of zero-mean normals with standard deviation \( \sigma_1, \sigma_2 \).

We know that the pdf \( f_Z(z) \) of \( Z = X + Y \) is given by \( f_X * f_Y \)

\[
f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_X(z - x) \, dx
\]

\[
= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp \left\{ - \left( \frac{x^2}{2\sigma_1^2} + \frac{(z - x)^2}{2\sigma_2^2} \right) \right\} \, dx
\]

Notice that the exponent is a quadratic in \( x \). To solve this we will complete the square and write it in the form

\[
\exp \left\{ - \left( \frac{(x - \mu')^2}{2\sigma'^2} + c \right) \right\}
\]

Why? This integral can be compared to the integral of a Gaussian pdf (which we know is 1). This is a very useful strategy in general with integrals resembling the Gaussian.

\[
\int_{-\infty}^{\infty} \exp \left\{ - \left( \frac{(x - \mu')^2}{2\sigma'^2} + c \right) \right\} \, dx = e^{-c} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma'}} \exp \left\{ - \frac{(x - \mu')^2}{2\sigma'^2} \right\} \, dx
\]

\[
= \frac{\sigma'e^{-c}}{\sqrt{2\pi\sigma'_1\sigma'_2}}
\]

So the final answer will be

\[
\frac{\sigma' e^{-c}}{\sqrt{2\pi\sigma'_1\sigma'_2}}
\]

We find \( \mu', \sigma' \) and \( c \) as follows -

\[
\frac{x^2}{2\sigma_1^2} + \frac{z^2 - 2zx + x^2}{2\sigma_2^2} = \frac{1}{2} \left\{ x^2 \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) - 2 \frac{x}{\sigma_2^2} + \frac{x}{\sigma_2^2} \right\}
\]

\[
= \frac{1}{2} \left\{ (ax)^2 + 2a\frac{b}{a}x + \left( \frac{b}{a} \right)^2 - \left( \frac{b}{a} \right)^2 + \frac{z^2}{\sigma_2^2} \right\}
\]

\[
= \frac{1}{2a^2} \left( x + \frac{b}{a^2} \right)^2 + \frac{1}{2} \left( \frac{z^2}{\sigma_2^2} - \frac{b^2}{a^2} \right)
\]

We then see that

\[
\sigma' = \frac{1}{a} = \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}
\]
and
\[ c = \frac{1}{2} \left\{ \frac{z^2}{\sigma_1^2} - \frac{z^2}{\sigma_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \right\} = \frac{z^2}{2\sigma_2^2} \left( 1 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right) = \frac{z^2}{2(\sigma_1^2 + \sigma_2^2)} \]

Plugging this back into (1) gives us the answer in all its glory -

\[ f_Z(z) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp \left\{ -\frac{z^2}{2(\sigma_1^2 + \sigma_2^2)} \right\} \]

Phew! This shows that the sum of two independent Gaussians is Gaussian and the variance as we’d expect from linearity is \( \sigma_1^2 + \sigma_2^2 \).

(Note: If you used MGFs or followed the procedure with a few calculation mistakes you may still give yourself full credit.)

(b) If \( Z \sim \mathcal{N}(0, 1) \), then \( \mathbb{E}[X^n] = \sigma^n \mathbb{E}[Z^n] \) so it suffices to work with \( \mathbb{E}[Z^n] \).

For odd \( n \), \( \mathbb{E}[Z^n] = 0 \) because the density of the Gaussian is symmetric around 0. For even \( n \geq 2 \), by integration by parts,
\[
\mathbb{E}[Z^n] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^n \exp \left( -\frac{z^2}{2} \right) dz = \frac{n-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{n-2} \exp \left( -\frac{z^2}{2} \right) dz = (n-1) \mathbb{E}[Z^{n-2}].
\]

Iterating this, we find that \( \mathbb{E}[Z^n] = (n-1)!! = \prod_{i=1}^{n/2} (2i-1) \). Another way to write this is \( \mathbb{E}[Z^n] = (n-1)!! = \frac{n!}{2 \times 4 \times \cdots \times (n-2) \times n} = \frac{n!}{2^{n/2}(n/2)!} \).

(c) Let \( W := (Z, \mathbb{1}\{Z > c\}) \). \( \mathbb{E}[Z] = 0 \) and \( \mathbb{E}[\mathbb{1}\{Z > c\}] = \mathbb{P}(Z > c) = \Phi(-c) \) by symmetry of the Gaussian; also, \( \text{var} Z = 1 \) and (using the variance of an indicator) \( \text{var} \mathbb{1}\{Z > c\} = \Phi(-c)(1 - \Phi(-c)) \). It remains to compute
\[
\text{cov}(Z, \mathbb{1}\{Z > c\}) = \mathbb{E}[Z \mathbb{1}\{Z > c\}] = \int_{c}^{\infty} z\phi(z) \, dz = \phi(c),
\]

where we have used the fact that \( \phi'(z) = -z\phi(z) \). Hence,
\[
\mathbb{E}[W] = \begin{bmatrix} 0 \\ \Phi(-c) \end{bmatrix}, \quad \text{cov} W = \begin{bmatrix} 1 & \phi(c) \\ \phi(c) & \Phi(-c)(1 - \Phi(-c)) \end{bmatrix}.
\]

4. Joint Density for Exponential Distribution

(a) If \( X \sim \text{Exp}(\lambda) \) and \( Y \sim \text{Exp}(\mu) \), \( X \) and \( Y \) independent, compute \( \mathbb{P}(X < Y) \).

(b) If \( X_k, 1 \leq k \leq n \) are exponentially distributed with parameters \( \lambda_1, \ldots, \lambda_n \), show that,
\[
\mathbb{P}(X_i = \min_{1 \leq k \leq n} X_k) = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}
\]

Solution:
(a) 

\[ P(X < Y) = \int_{y=0}^{\infty} P(X < y | Y = y) f_Y(y) dy \]

Since \( X \) and \( Y \) are independent, \( P(X < y | Y = y) = P(X < y) \), and since \( X \sim \text{Exp}(\lambda) \) and \( Y \sim \text{Exp}(\mu) \), \( P(X < y) = 1 - e^{-\lambda y} \) and \( f_Y(y) = \mu e^{-\mu y} \). Plugging in, we get, \( P(X < Y) = \frac{\lambda}{\lambda + \mu} \).

(b) We need to verify a nice fact about a collection of independent exponentially distributed random variable. Given a collection of random variables, \( Y_i \sim \text{Exp}(\mu_i), 1 \leq i \leq n \), \( \min(Y_i, 1 \leq i \leq n) \) is exponentially distributed with parameter \( \sum_{i=1}^{n} \mu_i \). This can be easily checked by considering the cdf of \( \min(Y_i) \). (Try it out!)

Now, \( P(X_i = \min_{1 \leq k \leq n} X_k) = P(X_i \leq \min_{1 \leq k \leq n, k \neq i} X_k) \). From the previous argument, \( \min_{1 \leq k \leq n, k \neq i} X_k \sim \sum_{j=1, j \neq i}^{n} \lambda_j \). Using the result of part (a), the claim follows.

5. Matrix Sketching

Matrix sketching is an important technique in randomized linear algebra to do large computations efficiently. For example, to compute the multiplication \( A^T \times B \) of two large matrices \( A \) and \( B \), we can use a random sketch matrix \( S \) to compute a "sketch" \( SA \) of \( A \) and a "sketch" \( SB \) of \( B \). Such a sketching matrix has the property that \( S^T S \approx I \) so that the approximate multiplication \( A^T S^T SB \) is close to \( A^T B \).

In this problem, we will discuss two popular sketching schemes and understand how they help in approximate computation. Let \( \hat{I} = S^T S \) and the dimension of sketch matrix \( S \) be \( d \times n \) (typically \( d \ll n \)).

(a) (Gaussian-sketch) Define

\[
S = \frac{1}{\sqrt{d}} \begin{bmatrix}
S_{11} & \cdots & S_{1n} \\
\vdots & \ddots & \vdots \\
S_{d1} & \cdots & S_{dn}
\end{bmatrix}
\]

such that \( S_{ij} \)'s are chosen i.i.d. from \( \mathcal{N}(0, 1) \) for all \( i \in [1, d] \) and \( j \in [1, n] \). Find the element-wise mean and variance (as a function of \( d \)) of the matrix \( \hat{I} = S^T S \), that is, find \( \mathbb{E}[\hat{I}_{ij}] \) and \( \text{Var}[\hat{I}_{ij}] \) for all \( i \in [1, n] \) and \( j \in [1, n] \).

(b) (Count-sketch) For each column \( j \in [1, n] \) of \( S \), choose a row \( i \) uniformly randomly from \( [1, d] \) such that

\[
S_{ij} = \begin{cases} 
1, & \text{with probability 0.5} \\
-1, & \text{with probability 0.5}
\end{cases}
\]

and assign \( S_{kj} = 0 \) for all \( k \neq i \). An example of a \( 3 \times 8 \) count-sketch is

\[
S = \begin{bmatrix}
0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\
1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & -1
\end{bmatrix}
\]

Again, find the element-wise mean and variance (as a function of \( d \)) of the matrix \( \hat{I} = S^T S \).
Note that for sufficiently large $d$, the matrix $\hat{I}$ is close to the identity matrix for both cases. We will use this fact in the lab to do an approximate matrix multiplication.

**Note:** You can use the fact that the fourth moment of a standard Gaussian is 3 without proof.

**Solution:** Let $\hat{I} = S^T S$.

(a) For the Gaussian-sketch $\hat{I}_{ij} = \frac{1}{d} \sum_{k=1}^{d} S_{ki} S_{kj}$. Thus, by using linearity of expectation and the fact that $S_{ki}$'s are drawn i.i.d. from $\mathcal{N}(0, 1)$, we get

$$E[\hat{I}_{ij}] = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases}$$

By the definition of variance, we have

$$\text{Var}[\hat{I}_{ij}] = \frac{1}{d^2} \left( E\left[ \left( \sum_{k=1}^{d} S_{ki} S_{kj} \right)^2 \right] - \left( E\left[ \sum_{k=1}^{d} S_{ki} S_{kj} \right] \right)^2 \right)$$

$$= \frac{1}{d^2} \left( E\left[ \left( \sum_{k=1}^{d} S_{ki} S_{kj} \right)^2 \right] - \left( \sum_{k=1}^{d} E[S_{ki} S_{kj}] \right)^2 \right)$$

Next, we consider two cases when $i = j$ and when $i \neq j$. When $i = j$

$$\text{Var}[\hat{I}_{ii}] = \frac{1}{d^2} \left( \sum_{k=1}^{d} E[S_{ki}^4] + \sum_{k=1, l=1}^{d} E[S_{ki}^2 E[S_{ki}^2] - d^2 \right)$$

$$= \frac{1}{d^2} \left( \sum_{k=1}^{d} E[S_{ki}^4] + 0 - d^2 \right)$$

$$= \frac{1}{d^2} \left( 3d + d(d - 1) - d^2 \right) = \frac{2}{d}$$

where we use the fact that the fourth moment of a standard Gaussian random variable is 3.

For the case when $i \neq j$, we use the fact that $S_{ki}$ and $S_{kj}$ are independent and get

$$\text{Var}[\hat{I}_{ij}] = \frac{1}{d^2} \left( E\left[ \left( \sum_{k=1}^{d} S_{ki} S_{kj} \right)^2 \right] - \left( \sum_{k=1}^{d} E[S_{ki}] E[S_{kj}] \right)^2 \right)$$

$$= \frac{1}{d^2} \left( \sum_{k=1}^{d} E[S_{ki}^2 E[S_{kj}^2] + \sum_{k=1, l=1}^{d} E[S_{ki}] E[S_{kj}] E[S_{ki}] E[S_{kj}] - 0 \right)$$

$$= \frac{1}{d^2} (d + 0) = \frac{1}{d}.$$
Thus, we have
\[
\text{Var}[\hat{I}_{ij}] = \begin{cases} 
2/d, & \text{if } i = j \\
1/d, & \text{otherwise.}
\end{cases}
\]

(b) Note that for Count-sketch, we have \( \hat{I}_{ij} = \sum_{k=1}^{d} S_{ki}S_{kj} \). By construction of \( S \), the diagonal terms \( \hat{I}_{ii} \) are always one. Thus, we only need to worry about the non-diagonal terms. It is also important to note that in \( S \), entries in a row are independent but the entries in a column are dependent (there can only be one non-zero entry in one column, as shown in the example). Also,
\[
S_{ki}S_{kj} = \begin{cases} 
1, & \text{with probability } 1/2d \\
-1, & \text{with probability } 1/2d \\
0, & \text{with probability } 1 - 1/d.
\end{cases} \quad \forall \ i \neq j.
\]
Thus,
\[
\mathbb{E}[\hat{I}_{ij}] = \begin{cases} 
1, & \text{if } i = j \\
0, & \text{otherwise.}
\end{cases}
\]
The diagonal terms in \( \hat{I} \) are exactly one, and hence, their variance is zero. For the non-diagonal terms, i.e. when \( i \neq j \), we have
\[
\text{Var}[\hat{I}_{ij}] = \mathbb{E}\left[ \left( \sum_{k=1}^{d} S_{ki}S_{kj} \right)^2 \right] - \left( \mathbb{E}\left[ \sum_{k=1}^{d} S_{ki}S_{kj} \right] \right)^2 \\
= \sum_{k=1}^{d} \mathbb{E}[S_{ki}^2]\mathbb{E}[S_{kj}^2] + \sum_{k=1, l=1, k \neq l}^{d} \mathbb{E}[S_{ki}S_{li}]\mathbb{E}[S_{kj}S_{lj}] - 0 \\
= \sum_{k=1}^{d} \frac{1}{d^2} + 0 = \frac{1}{d}.
\]
where the 0 in the last step comes from the fact at in any column \( j \), the product of two elements \( S_{kj}, S_{lj} \) is 0 since only one can be non-zero. Hence, the element-wise variance is
\[
\text{Var}[\hat{I}_{ij}] = \begin{cases} 
0, & \text{if } i = j \\
1/d, & \text{otherwise.}
\end{cases}
\]

6. **Records** Let \( n \) be a positive integer and \( X_1, X_2, \ldots, X_n \) be a sequence of i.i.d. continuous random variable with common probability density \( f_X \). For any integer \( 2 \leq k \leq n \), define \( X_k \) as a record-to-date of the sequence if \( X_k > X_i \) for all \( i = 1, \ldots, k-1 \). (\( X_1 \) is automatically a record-to-date.)

(a) Find the probability that \( X_2 \) is a record-to-date.

Hint: You should be able to do it without rigorous computation.

(b) Find the probability that \( X_n \) is a record-to-date.
(c) Find the expected number of records-to-date that occur over the first $n$ trials (Hint: Use indicator functions.) Compute this when $n \to \infty$.

Solution:

(a) $X_2$ is record-to-date with probability $1/2$. The reason is that $X_1$ and $X_2$ are i.i.d., so either one is larger than other with probability $1/2$. This uses the fact that they are equal with probability 0, since they have a density.

(b) Now, by the same symmetry argument, each $X_i$ for $i = 1, \ldots, n$ is equally likely to the the largest, so that each is largest with probability $1/n$. Since $X_n$ is the record-to-date if it is the largest among $X_1, \ldots, X_n$, it is a record with probability $1/n$.

**Remark**: An incorrect argument proceeds as follows. Since $X_1, \ldots, X_n$ are i.i.d., the probability of the event $\{X_n \geq X_i\}$ is $1/2$ for all $i = 1, \ldots, n$, and so the probability that $X_n$ is a record is $2^{-(n-1)}$ by independence. The reason why this argument is incorrect is because the events $\{X_n \geq X_i\}$ are not independent for different $i$; indeed, conditioned on the event $\{X_n \geq X_1\}$, it is reasonable to think that $X_n$ is probably large, which increases the probability of the event $\{X_n \geq X_2\}$.

(c) For $i = 1, \ldots, n$, let $\mathbb{1}_i$ be 1 if $X_i$ is a record-to-date, 0 otherwise. Thus $\mathbb{E}(\mathbb{1}_i)$ is the expected value of the number of records-to-date on trial $i$. Thus,

$$\mathbb{E}(\mathbb{1}_i) = \mathbb{P}(\mathbb{1}_i = 1) = \frac{1}{i}.$$ 

Thus,

$$\mathbb{E}(\text{records to date in } n \text{ trials}) = \sum_{i=1}^{n} \mathbb{E}(\mathbb{1}_i) = \sum_{i=1}^{n} \frac{1}{i}.$$ 

This is a harmonic series, and if $n \to \infty$, it diverges to $\infty$. 