UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Homework 04

Fall 2023

1. **Basic Properties of Jointly Gaussian Random Variables**

Let $(X_1, \dots, X_n)$ be a collection of jointly Gaussian random variables with mean vector μ and covariance matrix Σ. Their joint density is given by, for $x \in \mathbb{R}^n$,

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left\{-\frac{1}{2}(x-\mu)^\mathsf{T}\Sigma^{-1}(x-\mu)\right\}.$$

a. Show that $X_1, \dots, X_n$ are independent if and only if they are pairwise uncorrelated.

b. Show that any linear combination of $X_1, \dots, X_n$ will also be a Gaussian random variable. *Hint*: Consider using moment-generating functions.

Solution:

a. Independence implies uncorrelatedness in general, so suppose $X_1, \dots, X_n$ are pairwise uncorrelated, in which case Σ is a diagonal matrix with diagonal entries $\sigma_1^2, \dots, \sigma_n^2$. Then

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{(2\pi)^n \prod_{i=1}^n \sigma_i^2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n \frac{1}{\sigma_i^2}(x_i-\mu_i)^2\right\} \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{1}{2\sigma_i^2}(x_i-\mu_i)^2\right\} \\ &= \prod_{i=1}^n f_{X_i}(x_i), \end{aligned}$$

so pairwise uncorrelatedness implies the independence of $X_1, \dots, X_n$ jointly Gaussian.

b. The moment-generating function of a linear combination $Y = a^\mathsf{T} X = \sum_{i=1}^n a_i X_i$ is

$$\begin{aligned} \phi_Y(t) &= \mathbb{E}(\exp\{ta^\mathsf{T}X\}) = \phi_X(ta) \\ &= \exp\left\{ta^\mathsf{T}\mu - \frac{1}{2}t^2 a^\mathsf{T}\Sigma a\right\}. \end{aligned}$$

Therefore Y is Gaussian with distribution $\mathcal{N}(a^\mathsf{T}\mu, a^\mathsf{T}\Sigma a)$.

2. **Gaussian Sine**

Let X, Y, Z be jointly Gaussian random variables with covariance matrix

$$\begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix}$$

and mean vector $[0, 2, 0]$. Compute $\mathbb{E}[(\sin X)Y(\sin Z)]$. *Hint:* Condition on (X, Z).

Solution: Conditioning on (X, Z), we have by tower property that

$$\mathbb{E}[\sin(X)Y\sin(Z)] = \mathbb{E}[\mathbb{E}[\sin(X)Y\sin(Z)|X, Z]] = \mathbb{E}[\sin(X)\sin(Z)\,\mathbb{E}[Y|X, Z]]. \tag{1}$$

The inner conditional expectation can be computed as

$$\begin{aligned} \mathbb{E}[Y|X, Z] &= \mu_Y + \Sigma_{Y,(X,Z)}\Sigma_{(X,Z)}^{-1}\begin{bmatrix} X \\ Z \end{bmatrix} \\ &= 2 + [1\ 1]\begin{bmatrix} 1/4 & 0 \\ 0 & 1/4 \end{bmatrix}\begin{bmatrix} X \\ Z \end{bmatrix} \\ &= 2 + \frac{1}{4}X + \frac{1}{4}Z. \end{aligned}$$

Then (1) becomes (we use the fact that X and Z are uncorrelated and therefore independent, and since sin is odd and X and Z are symmetric about 0, $\mathbb{E}[\sin(X)] = \mathbb{E}[\sin(Z)] = 0$):

$$\begin{aligned} &\mathbb{E}[\sin(X)\sin(Z)(2 + \frac{1}{4}X + \frac{1}{4}Z)] \\ &= 2\,\mathbb{E}[\sin(X)]\,\mathbb{E}[\sin(Z)] + \mathbb{E}[\sin(X)X/4]\,\mathbb{E}[\sin(Z)] + \mathbb{E}[\sin(X)]\,\mathbb{E}[\sin(Z)Z/4] \\ &= 0. \end{aligned}$$

3. **Lognormal Distribution and the Moment Problem**

(Optional) This question seeks to answer the following question: if two distributions have the same moments of all orders, are they necessarily the same? An equivalent way to phrase the problem is: if the moments exist, do they completely determine the distribution?

a. Suppose that Z is a standard Gaussian and let $X = e^Z$. Calculate the density of X. (This is known as the **lognormal** distribution.)

b. Let $f_X(x)$ denote the density of the lognormal density. Define

$$f_a(x) = f_X(x)(1 + a\sin(2\pi \log x)), \qquad x > 0, \quad -1 \le a \le 1.$$

Argue that $f_a(x)$ is a valid density function and show that $f_X(x)$ and $f_a(x)$ have the same moments of all orders by showing that

$$\int_0^\infty x^k f_X(x) \sin(2\pi \log x)\,\mathrm{d}x = 0, \qquad k \in \mathbb{N}.$$

c. Explicitly calculate the moments of the lognormal distribution.

d. Now, let Y_b $(b > 0)$ be a discrete random variable with distribution

$$\Pr(Y_b = be^n) = cb^{-n}e^{-n^2/2}, \qquad n \in \mathbb{Z},$$

where c is chosen to normalize the distribution:

$$\sum_{n=-\infty}^{\infty} cb^{-n}e^{-n^2/2} = 1.$$

Show that Y_b has the same moments as X. This provides a discrete counterexample to the moment problem.

Solution:

a. Observe that

$$\Pr(X \le x) = \Pr(Z \le \log x) = \Phi(\log x),$$

for $x > 0$. Differentiating the CDF, we have

$$f_X(x) = \phi(\log x) \cdot \frac{1}{x} = \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2}, \qquad x > 0.$$

b. The goal is to prove

$$\int_0^\infty x^k \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2} \sin(2\pi \log x)\,\mathrm{d}x = 0.$$

Change variables with $x = e^{s+k}$, $s = \log x - k$, and $\mathrm{d}s = x^{-1}\,\mathrm{d}x$. One has

$$\int_{-\infty}^\infty e^{ks+k^2} e^{-(s+k)^2/2} \sin(2\pi(s+k))\,\mathrm{d}s = e^{k^2/2} \int_{-\infty}^\infty e^{-s^2/2} \sin(2\pi s)\,\mathrm{d}s,$$

where we used the fact that $\sin(x)$ is 2π-periodic and k is an integer. Now, we observe that the integrand is an odd function, so the integral is 0 as desired. Hence, $f_X(x)$ and $f_a(x)$ have the same moments of all orders. For the case of $k = 0$, we see that $f_X(x)$ and $f_a(x)$ both integrate to 1, and the condition $-1 \le a \le 1$ ensures that $f_a(x) \ge 0$, so $f_a(x)$ is a valid density function.

This provides a negative answer to the moment problem.

c. The moments are

$$\mu_k = E[X^k] = E[e^{kZ}] = \int_{-\infty}^{\infty} e^{kz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, \mathrm{d}z$$
$$= e^{k^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-k)^2/2}\, \mathrm{d}z = e^{k^2/2}.$$

(The calculation above essentially calculates the moment-generating function of Z.)

d. One has

$$E[Y_b^k] = \sum_{n=-\infty}^{\infty} (be^n)^k c b^{-n} e^{-n^2/2} = e^{k^2/2} \sum_{n=-\infty}^{\infty} c b^{-(n-k)} e^{-(n-k)^2/2}$$
$$= e^{k^2/2}.$$

Remark: Let ν_k be the kth **absolute moment**, that is, $\nu_k = E[|X|^k]$. A sufficient condition for the moments to uniquely determine the distribution is

$$\limsup_{k\to\infty} \frac{\nu_k^{1/k}}{k} < \infty.$$

4. **Revisiting Proofs Using Transforms**

   a. Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ be independent. Calculate the MGF of $X + Y$, and use this to show that $X + Y \sim \text{Poisson}(\lambda + \mu)$.

   b. Calculate the MGF of $X \sim \text{Exponential}(\lambda)$, and use this to find all of the moments of X.

   c. Repeat the above part, but for $X \sim \mathcal{N}(0, 1)$.

Solution:

a. The MGF of X is

$$\begin{aligned}\mathbb{E}(\exp(sX)) &= \sum_{x \in \mathbb{N}} \exp(sx) \frac{\lambda^x \exp(-\lambda)}{x!} \\ &= \exp(-\lambda) \sum_{x \in \mathbb{N}} \frac{(\lambda \exp s)^x}{x!} = \exp(\lambda(\exp s - 1)),\end{aligned}$$

which converges for all $s \in \mathbb{R}$. The MGF of $X + Y$ is

$$\begin{aligned}\mathbb{E}(\exp(s(X+Y))) = \mathbb{E}(\exp(sX) \cdot \exp(sY)) &= \mathbb{E}(\exp(sX)) \cdot \mathbb{E}(\exp(sY)) \\ &= \exp(\lambda(\exp s - 1)) \cdot \exp(\mu(\exp s - 1)) \\ &= \exp((\lambda + \mu)(\exp s - 1)),\end{aligned}$$

which we recognize as the MGF of a $\text{Poisson}(\lambda + \mu)$ random variable.

Remark: In general, it is not easy to argue that the MGF uniquely determines the probability distribution, which requires a few assumptions on the MGF itself, but we will not worry about these issues in this course.

b. We calculate

$$M_X(s) = \int_0^\infty \exp(sx) \lambda \exp(-\lambda x)\, dx = \lambda \int_0^\infty \exp(-(\lambda - s)x)\, dx = \frac{\lambda}{\lambda - s},$$

which converges for $s < \lambda$. Expanding M_X as a geometric series,

$$M_X(s) = \frac{1}{1 - \frac{s}{\lambda}} = \sum_{k \in \mathbb{N}} \left(\frac{s}{\lambda}\right)^k,$$

as long as $|s| < \lambda$. Comparing the last expression with

$$\mathbb{E}(\exp(sX)) = \mathbb{E}\left(\sum_{k \in \mathbb{N}} \frac{(sX)^k}{k!}\right) = \sum_{k \in \mathbb{N}} \frac{s^k\, \mathbb{E}(X^k)}{k!}$$

and matching terms, we can argue that $\mathbb{E}(X^k) = \frac{k!}{\lambda^k}$.

c. The MGF of the standard Gaussian is

$$\begin{aligned}M_X(s) &= \int_{-\infty}^\infty \exp(sx) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left(-\frac{1}{2}(x^2 - 2sx)\right) dx\end{aligned}$$

$$
\begin{aligned}
&= \exp\left(\frac{s^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-s)^2}{2}\right)\, dx \\
&= \exp\left(\frac{s^2}{2}\right).
\end{aligned}
$$

Expanding as a power series,

$$M_X(s) = \sum_{k \in \mathbb{N}} \frac{s^{2k}}{2^k k!},$$

so by comparing terms as before, we see that $\mathbb{E}(X^k) = 0$ if k is odd, and

$$\mathbb{E}(X^k) = \frac{k!}{2^{k/2}(\frac{k}{2})!} = (k-1)!!$$

if k is even.

5. **Coupon Collector Bounds**

Recall the coupon collector's problem, in which there are n different types of coupons. Every box contains a single coupon, and we let the random variable X be the number of boxes bought until one of every type of coupon is obtained. The expected value of X is nH_n, where $H_n := \sum_{i=1}^{n} \frac{1}{i}$ is the *harmonic number* of order n, which satisfies the inequality

$$\ln n \leq H_n \leq \ln n + 1.$$

a. Use Markov's inequality in order to show that

$$\mathbb{P}(X > 2nH_n) \leq \frac{1}{2}.$$

b. Use Chebyshev's inequality in order to show that

$$\mathbb{P}(X > 2nH_n) \leq \frac{\pi^2}{6(\ln n)^2}.$$

Note: You can use Euler's solution to the Basel problem, the identity $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$.

c. Define appropriate events and use the union bound in order to show that

$$\mathbb{P}(X > 2nH_n) \leq \frac{1}{n}.$$

Note: $a_n = (1 - \frac{1}{n})^n$ is a strictly increasing sequence with limit e^{-1}.

Solution:

a. We are given $\mathbb{E}(X) = nH_n$, so

$$\mathbb{P}(X > 2nH_n) \leq \frac{\mathbb{E}(X)}{2nH_n} = \frac{1}{2}.$$

b. We can write X as an independent sum $\sum_{i=1}^{n} X_i$, where $X_i \sim \text{Geometric}(\frac{n-i+1}{n})$, so

$$\text{var}(X) = \sum_{i=1}^{n} \text{var}(X_i) < \sum_{i=1}^{n} \left(\frac{n}{n-i+1}\right)^2 = \sum_{i=1}^{n} \left(\frac{n}{i}\right)^2 < \frac{\pi^2 n^2}{6}.$$

Using Chebyshev's inequality, we have that

$$\mathbb{P}(X > 2nH_n) \leq \mathbb{P}(|X - nH_n| > nH_n) \leq \frac{\text{var}(X)}{(nH_n)^2} < \frac{\pi^2}{6H_n^2} \leq \frac{\pi^2}{6(\ln n)^2}.$$

c. Let A_i be the event that we fail to get box i after $2nH_n$ tries.

$$\mathbb{P}(A_i) \leq \left(\frac{n-1}{n}\right)^{2nH_n} = \left[\left(1 - \frac{1}{n}\right)^n\right]^{2H_n} < e^{-2H_n} \leq e^{-2\ln n} = \frac{1}{n^2}.$$

Now, by the union bound, we can conclude that

$$\mathbb{P}(X > 2nH_n) = \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i) \leq \frac{1}{n}.$$

6. **Matrix Sketching**

Matrix sketching is an important technique in randomized linear algebra for doing large computations efficiently. For example, to compute $\mathbf{A}^T \times \mathbf{B}$ for two large matrices $\mathbf{A}$ and $\mathbf{B}$, we can use a random sketch matrix $\mathbf{S}$ to compute a "sketch" $\mathbf{SA}$ of $\mathbf{A}$, and a sketch $\mathbf{SB}$ of $\mathbf{B}$. Such a sketching matrix has the property that

$$\mathbf{S}^T\mathbf{S} \approx \mathbf{I},$$

so that the approximate multiplication $(\mathbf{SA})^T(\mathbf{SB}) = \mathbf{A}^T\mathbf{S}^T\mathbf{SB}$ is close to $\mathbf{A}^T\mathbf{B}$.

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

a. **Gaussian sketch**. Let the sketch matrix be

$$\mathbf{S} = \frac{1}{\sqrt{d}} \begin{bmatrix} S_{1,1} & \cdots & S_{1,n} \\ \vdots & \ddots & \vdots \\ S_{d,1} & \cdots & S_{d,n} \end{bmatrix},$$

where the $S_{i,j}$ are chosen i.i.d. from $\mathcal{N}(0,1)$ for all $i \in [1,d]$ and $j \in [1,n]$. Show that the elementwise mean and variance of the matrix $\hat{\mathbf{I}}$, as functions of d, are

$$\mathbb{E}(\hat{I}_{i,j}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\text{var}(\hat{I}_{i,j}) = \begin{cases} \frac{2}{d} & \text{if } i = j \\ \frac{1}{d} & \text{otherwise.} \end{cases}$$

You can use without proof the fact that $\mathbb{E}(Z^4) = 3$ for $Z \sim \mathcal{N}(0,1)$.

b. **Count sketch**. For each column $j \in [1,n]$ of $\mathbf{S}$, choose a row i uniformly randomly from $[1,d]$. Set

$$S_{i,j} = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2}, \end{cases}$$

and assign $S_{k,j} = 0$ for all $k \neq i$. An example of a 3×8 count sketch matrix is

$$\begin{bmatrix} 0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Show that the elementwise mean and variance of the matrix $\hat{\mathbf{I}}$ are

$$\mathbb{E}(\hat{I}_{i,j}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\text{var}(\hat{I}_{i,j}) = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{d} & \text{otherwise.} \end{cases}$$

Note that for sufficiently large d, the matrix $\hat{\mathbf{I}}$ is close to the identity matrix in both cases. We use this fact in the lab to do an approximate matrix multiplication.

Solution:

a. For the Gaussian sketch matrix $\mathbf{S}$, we have

$$\hat{I}_{i,j} = \frac{1}{d} \sum_{k=1}^{d} S_{k,i} S_{k,j}.$$

By the linearity of expectation, and the $S_{k,i}$ being drawn i.i.d. from $\mathcal{N}(0,1)$, we get

$$\mathbb{E}(\hat{I}_{i,j}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Then, by the definition of variance, we have

$$\begin{aligned} d^2 \operatorname{var}(\hat{I}_{i,j}) &= \mathbb{E}[(d\hat{I}_{i,j})^2] - \mathbb{E}[d\hat{I}_{i,j}]^2 \\ &= \mathbb{E}\left[\left(\sum_{k=1}^{d} S_{k,i} S_{k,j}\right)^2\right] - d^2 \, \mathbb{1}_{i=j} \,. \end{aligned}$$

Now we consider the two cases of $i = j$ and $i \neq j$, starting with the former:

$$\begin{aligned} d^2 \operatorname{var}(\hat{I}_{i,i}) &= \sum_{k=1}^{d} \mathbb{E}(S_{k,i}^4) + \sum_{k \neq \ell} \mathbb{E}(S_{k,i}^2)\, \mathbb{E}(S_{\ell,i}^2) - d^2 \\ &= 3d + d(d-1) - d^2 = 2d. \end{aligned}$$

For the case of $i \neq j$, we can use the independence of $S_{k,i}$ and $S_{k,j}$:

$$\begin{aligned} d^2 \operatorname{var}(\hat{I}_{i,j}) &= \sum_{k=1}^{d} \mathbb{E}(S_{k,i}^2)\, \mathbb{E}(S_{k,j}^2) + \sum_{k \neq \ell} \mathbb{E}(S_{k,i})\, \mathbb{E}(S_{k,j})\, \mathbb{E}(S_{\ell,i})\, \mathbb{E}(S_{\ell,j}) \\ &= d + 0 = d. \end{aligned}$$

Thus the elementwise variance is

$$\operatorname{var}(\hat{I}_{i,j}) = \begin{cases} \frac{2}{d} & \text{if } i = j \\ \frac{1}{d} & \text{otherwise.} \end{cases}$$

b. For the count sketch matrix $\mathbf{S}$, we have

$$\hat{I}_{i,j} = \sum_{k=1}^{d} S_{k,i} S_{k,j}.$$

By construction of $\mathbf{S}$, the diagonal terms $\hat{I}_{i,i}$ are always 1, so their mean is 1 and their variance is 0, and we only need to worry about the non-diagonal terms.

We also note that in $\mathbf{S}$, entries in a row are independent, but entries in a column are dependent. (There can only be one nonzero entry in one column.) Moreover, for all $i \neq j$,

$$S_{k,i}S_{k,j} = \begin{cases} 1 & \text{with probability } \frac{1}{2d^2} \\ -1 & \text{with probability } \frac{1}{2d^2} \\ 0 & \text{with probability } 1 - \frac{1}{d^2}. \end{cases}$$

Thus the elementwise expectation is

$$\mathbb{E}(\hat{I}_{i,j}) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Now, for $i \neq j$, using the fact that $\mathbb{E}[\hat{I}_{i,j}]^2 = 0$,

$$\begin{aligned} \mathrm{var}(\hat{I}_{i,j}) &= \mathbb{E}\left[\left(\sum_{k=1}^{d} S_{k,i}S_{k,j}\right)^2\right] \\ &= \sum_{k=1}^{d} \mathbb{E}(S_{k,i}^2)\,\mathbb{E}(S_{k,j}^2) + \sum_{k\neq\ell} \mathbb{E}(S_{k,i}S_{\ell,i})\,\mathbb{E}(S_{k,j}S_{\ell,j}) \\ &= \sum_{k=1}^{d} \frac{1}{d^2} + 0 = \frac{1}{d}. \end{aligned}$$

The term 0 in the last step comes from the fact that in any column j, the product of two elements $S_{k,j}S_{\ell,j} = 0$, since only one can be nonzero. Thus the elementwise variance is

$$\mathrm{var}(\hat{I}_{i,j}) = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{d} & \text{otherwise.} \end{cases}$$