Problem 1: Review of Random Variables

Let $X$ and $Y$ be discrete random variables defined on some probability space with a joint pmf $p_{XY}(x, y)$. Let $a, b \in \mathbb{R}$ be fixed.

(a) Prove that $E[aX + bY] = aE[X] + bE[Y]$. Do not assume independence.

(b) Prove that if $X$ and $Y$ are independent random variables, then $E[X \cdot Y] = E[X] \cdot E[Y]$.

(c) Assume that $X$ and $Y$ are not independent. Find an example where $E[X \cdot Y] \neq E[X] \cdot E[Y]$, and another example where $E[X \cdot Y] = E[X] \cdot E[Y]$.

(d) Prove that if $X$ and $Y$ are independent, then they are also uncorrelated, i.e.,

$$Cov(X, Y) := E[(X - E[X])(Y - E[Y])] = 0. \quad (1)$$

(e) Find an example where $X$ and $Y$ are uncorrelated but dependent.

(f) Assume that $X$ and $Y$ are uncorrelated and let $\sigma^2_X$ and $\sigma^2_Y$ be the variances of $X$ and $Y$, respectively. Find the variance of $aX + bY$ and express it in terms of $\sigma^2_X, \sigma^2_Y, a, b$.

Hint: First show that $Cov(X, Y) = E[X \cdot Y] - E[X] \cdot E[Y]$.

Solution 1. (a)

$$E[aX + bY] = \sum_x \sum_y (ax + by)p_{XY}(x, y)$$

$$= \sum_x ax \sum_y p_{XY}(x, y) + \sum_y by \sum_x p_{XY}(x, y)$$

$$= a \sum_x x p_X(x) + b \sum_y y p_Y(y)$$

$$= aE[X] + bE[Y].$$

(b) If $X$ and $Y$ are independent, we have $p_{XY}(x, y) = p_X(x)p_Y(y)$, then

$$E[X \cdot Y] = \sum_x \sum_y xp_{XY}(x, y)$$

$$= \sum_x \sum_y xp_X(x)p_Y(y)$$

$$= \sum_x xp_X(x) \sum_y p_Y(y)$$

$$= E[X] \cdot E[Y].$$
(c) For the first example, suppose \( Pr(X = 0, Y = 1) = Pr(X = 1, Y = 0) = \frac{1}{2} \), and \( Pr(X = 0, Y = 0) = Pr(X = 1, Y = 1) = 0 \). \( X, Y \) are dependent, and we have \( \mathbb{E}[X \cdot Y] = 0 \) while \( \mathbb{E}[X] \mathbb{E}[Y] = \frac{1}{4} \).

For the second example, suppose \( Pr(X = -1, Y = 0) = Pr(X = 0, Y = 1) = Pr(X = 1, Y = 0) = \frac{1}{4} \). \( X, Y \) are dependent. Obviously we have \( \mathbb{E}[X \cdot Y] = 0 \), and furthermore \( \mathbb{E}[X] = 0 \), hence \( \mathbb{E}[X] \mathbb{E}[Y] = 0 \).

(d) If \( X \) and \( Y \) are independent, we have \( p_{XY}(x, y) = p_X(x)p_Y(y) \), then
\[
\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \sum_x \sum_y (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) p_{XY}(x, y)
= \sum_x \sum_y (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) p_X(x)p_Y(y)
= \sum_x (x - \mathbb{E}[X]) p_X(x) \sum_y (y - \mathbb{E}[Y]) p_Y(y)
= (\mathbb{E}[X] - \mathbb{E}[X])(\mathbb{E}[Y] - \mathbb{E}[Y]) = 0.
\]

Thus, \( X \) and \( Y \) are uncorrelated.

(e) One example where \( X \) and \( Y \) are uncorrelated but dependent is
\[
P_{XY}(x, y) = \begin{cases} \frac{1}{4} & \text{if } (x, y) \in \{(-1, 0), (1, 0), (0, 1)\} \\ 0 & \text{otherwise.} \end{cases}
\]

First, it can be easily checked that \( \mathbb{E}[X \cdot Y] = 0 = \mathbb{E}[X] \cdot \mathbb{E}[Y] \) (note that \( \mathbb{E}[X] = 0 \)). Second, \( X \) and \( Y \) are dependent since \( P_{XY}(1, 0) = \frac{1}{4} \) but \( P_X(1)P_Y(0) = \frac{1}{3} \times \frac{4}{3} \).

(f) First, we have
\[
\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]
= \mathbb{E}[XY - X\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]]
= \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y].
\]

Thus, \( \text{Cov}(X, Y) = 0 \) if and only if \( \mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \).

Then,
\[
\sigma_{aX+bY}^2 = \mathbb{E}[(aX + bY - \mathbb{E}[aX + bY])^2
= \mathbb{E}[(aX + bY)^2] - (\mathbb{E}[aX + bY])^2
= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X \cdot Y] + b^2\mathbb{E}[Y^2] - a^2\mathbb{E}[X]^2 - 2ab\mathbb{E}[X]\mathbb{E}[Y] - b^2\mathbb{E}[Y]^2
= a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) + b^2(\mathbb{E}[Y^2] - \mathbb{E}[Y]^2)
= a^2\sigma_X^2 + b^2\sigma_Y^2.
\]

We remark that since the independence of \( X \) and \( Y \) implies \( \text{Cov}(X, Y) = 0 \), we also have \( \sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 \) if \( X \) and \( Y \) are independent.

**Problem 2: Review of Gaussian Random Variables**

A random variable \( X \) with probability density function
\[
p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}
\]

is called a **Gaussian** random variable.

(a) Explicitly calculate the mean \( \mathbb{E}[X] \), the second moment \( \mathbb{E}[X^2] \), and the variance \( \text{Var}[X] \) of the random variable \( X \).
(b) Let us now consider events of the following kind:

\[ \Pr(X < \alpha). \tag{3} \]

Unfortunately for Gaussian random variables this cannot be calculated in closed form. Instead, we will rewrite it in terms of the standard Q-function:

\[ Q(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \, du \tag{4} \]

Express \( \Pr(X < \alpha) \) in terms of the Q-function and the parameters \( m \) and \( \sigma^2 \) of the Gaussian pdf.

Like we said, the Q-function cannot be calculated in closed form. Therefore, it is important to have bounds on the Q-function. In the next 3 subproblems, you derive the most important of these bounds, learning some very general and powerful tools along the way:

(c) Derive the Markov inequality, which says that for any non-negative random variable \( X \) and positive \( a \), we have

\[ \Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \tag{5} \]

(d) Use the Markov inequality to derive the Chernoff bound: the probability that a real random variable \( Z \) exceeds \( b \) is given by

\[ \Pr(Z \geq b) \leq \mathbb{E}[e^{s(Z-b)}], \quad s \geq 0. \tag{6} \]

(e) Use the Chernoff bound to show that

\[ Q(x) \leq e^{-\frac{x^2}{2}} \text{ for } x \geq 0. \tag{7} \]

Solution 2. (a) First,

\[ \mathbb{E}[X] = \int_{-\infty}^{\infty} xp_X(x) \, dx \]

\[ = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} xe^{-(x-m)^2 / 2\sigma^2} \, dx \]

\[ \overset{(*)}{=} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} ue^{-\frac{u^2}{2\sigma^2}} \, du + m \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} \, du \tag{8} \]

\[ \overset{(*)}{=} 0 + m \]

\[ = m, \]

where \((*)\) follows by a change of variable \( u = x - m \) and \((\dagger)\) follows since the first integrand in \((??)\) is an odd function and the second integrand in \((??)\) is a probability density function. We remark that the integral

\[ \int_{-\infty}^{\infty} e^{-x^2} \, dx \]
known as Gaussian integral, can be evaluated explicitly to be $\sqrt{\pi}$. Second,

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 p_X(x) \, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 e^{-\frac{(x-m)^2}{2\sigma^2}} \, dx$$

$$\equiv \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2\sigma^2}} \, du + \frac{2m}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} u e^{-\frac{u^2}{2\sigma^2}} \, du + m^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} \, du \quad (9)$$

$$\equiv \sigma^2 + 0 + m^2 = \sigma^2 + m^2,$$

where $(\ast)$ follows by a change of variable $u = x - m$ and $(\dag)$ follows from the same arguments in the evaluation of $\mathbb{E}[X]$ and an integration by parts to the first integral in $(9)$:

$$\int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2\sigma^2}} \, du = -\frac{\sigma^2}{\sqrt{2\pi\sigma^2}} \left( \frac{ue^{-\frac{u^2}{2\sigma^2}}}{\sigma^2} \right)_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-\frac{u^2}{2\sigma^2}} \, du = 0 + \sigma^2.$$

Therefore,

$$\text{Var}[X] = \mathbb{E}[X - \mathbb{E}[X]]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2 + m^2 - m^2 = \sigma^2.$$ 

\[b\]

$$P(X < \alpha) = \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \, dx$$

$$\equiv \int_{-\infty}^{\frac{\alpha-m}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \, du$$

$$= 1 - Q \left( \frac{\alpha-m}{\sigma} \right),$$

where $(\ast)$ follows by a change of variable $u = \frac{x-m}{\sigma}$.

\[c\]

$$\mathbb{E}[X] = \int_{0}^{\alpha} x p_X(x) \, dx + \int_{\alpha}^{\infty} x p_X(x) \, dx$$

$$\geq 0 + a \int_{\alpha}^{\infty} p_X(x) \, dx$$

$$= aP(X \geq a).$$

\[d\] Fix $s \geq 0$, then we have

$$P(Z \geq b) \leq P(s(Z-b) \geq 0)$$

$$= P(e^{s(Z-b)} \geq e^0)$$

$$\leq \mathbb{E}[e^{s(Z-b)}],$$

where $(\ast)$ follows from the Markov inequality.
(c) Let $X$ be a Gaussian random variable with mean zero and unit variance, then we have

$$Q(x) = \mathbb{P}(X \geq x) \leq \mathbb{E} \left[ e^{s(X-x)} \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{su} e^{-u^2/2} du = e^{-sx^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du = e^{-sx^2/2},$$

where $(s)$ follows from the Chernoff bound. In order to get the tightest bound, we need to minimize $-sx + s^2/2$ which gives $s = x$ and then the desired bound is established.

**Problem 3: Moment Generating Function**

In the class we had considered the logarithmic moment generating function

$$\phi(s) := \ln \mathbb{E}[\exp(sX)] = \sum_x p(x) \exp(sx)$$

of a real-valued random variable $X$ taking values on a finite set, and showed that $\phi'(s) = \mathbb{E}[X_s]$ where $X_s$ is a random variable taking the same values as $X$ but with probabilities $p_s(x) := p(x) \exp(sx) \exp(-\phi(s))$.

(a) Show that

$$\phi''(s) = \text{Var}(X_s) := \mathbb{E}[X_s^2] - \mathbb{E}[X]^2$$

and conclude that $\phi''(s) \geq 0$ and the inequality is strict except when $X$ is deterministic.

(b) Let $x_{\min} := \min\{x : p(x) > 0\}$ and $x_{\max} := \max\{x : p(x) > 0\}$ be the smallest and largest values $X$ takes. Show that

$$\lim_{s \to -\infty} \phi'(s) = x_{\min}, \quad \text{and} \quad \lim_{s \to -\infty} \phi'(s) = x_{\max}.$$

**Solution 3.** (a) As $\phi(s) := \ln \mathbb{E}[\exp(sX)]$, we have

$$\phi'(s) = \frac{\mathbb{E}[X \exp(sX)]}{\mathbb{E}[\exp(sX)]} = \mathbb{E}[X \exp(sX) \exp(-\phi(s))] = \mathbb{E}[X_s] \quad (10)$$

$$\phi''(s) = \frac{\mathbb{E}[X^2 \exp(sX)]}{\mathbb{E}[\exp(sX)]} - \frac{\mathbb{E}[X \exp(sX)\mathbb{E}[X \exp(sX)]]}{\mathbb{E}[\exp(sX)]^2} \quad (11)$$

The second term is $\mathbb{E}[X_s^2]$ and the first term equals $\sum_x x^2 \exp(sx)/\exp(\phi(s)) = \mathbb{E}[X_s^2]$. So $\phi''(s) = \text{Var}(X_s)$. Moreover, $\text{Var}(X_s) \geq 0$ with equality only when $X_s$ is deterministic. But $X_s$ is deterministic only when $X$ is.

(b) Observe that

$$\phi'(s) = \frac{\mathbb{E}[X \exp(sX)]}{\mathbb{E}[\exp(sX)]} = \frac{\mathbb{E}[X \exp(sX)\exp(-sx_{\max})]}{\mathbb{E}[\exp(sX)]\exp(-sx_{\max})} \quad (12)$$

$$\phi'(s) = \frac{\sum_x p(x)x \exp(-s(x_{\max} - x))}{\sum_x p(x)\exp(-s(x_{\max} - x))} \quad (13)$$

In the sums above, as $s \to \infty$, all terms vanish except the ones for $x = x_{\max}$. Hence we have

$$\lim_{s \to \infty} \phi'(s) = \frac{p(x_{\max})x_{\max}}{p(x_{\max})} = x_{\max} \quad (14)$$

Similarly, we can show that $\lim_{s \to -\infty} \phi'(s) = x_{\min}$.
Problem 4: Hoeffding’s Lemma

Prove Lemma 2.3 in the lecture notes. In other words, prove that if \( X \) is a zero-mean random variable taking values in \([a, b]\) then

\[
\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2}{2}((a-b)^2/4)}.
\]

Expressed differently, \( X \) is \([(a-b)^2/4]\)-subgaussian.

Hint: You can use the following steps to prove the lemma:

1. Let \( \lambda > 0 \). Let \( X \) be a random variable such that \( a \leq X \leq b \) and \( \mathbb{E}[X] = 0 \). By considering the convex function \( x \rightarrow e^{\lambda x} \), show that

\[
\mathbb{E}[e^{\lambda X}] \leq \frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b}.
\]

(15)

2. Let \( p = -a/(b-a) \) and \( h = \lambda(b-a) \). Verify that the right-hand side of (8) equals \( e^{L(h)} \) where

\[
L(h) = -hp + \log(1 - p + pe^h).
\]

3. By Taylor’s theorem, there exists \( \xi \in (0, h) \) such that

\[
L(h) = L(0) + hL'(0) + \frac{h^2}{2} L''(\xi).
\]

Show that \( L(h) \leq h^2/8 \) and hence \( \mathbb{E}[e^{\lambda X}] \leq e^{\lambda^2(b-a)^2/8} \).

Solution 4. Since \( e^{\lambda x} \) is convex in \( x \) we have for all \( a \leq x \leq b \),

\[
e^{\lambda x} \leq \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b}.
\]

If we take the expected value of this wrt \( X \) and recall that \( \mathbb{E}[X] = 0 \) then it follows that

\[
\mathbb{E}[e^{\lambda X}] \leq \frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b}.
\]

Consider the right-hand side. Note that we must have \( a < 0 \) and \( b > 0 \) since \( \mathbb{E}[X] = 0 \). Set \( p = -a/(b-a) \), \( 0 \leq p \leq 1 \), and \( \lambda' = \lambda(b-a) \). The right-hand side can then be written as

\[
(1-p)e^{-\lambda' p} + pe^{\lambda'(1-p)} \leq e^{\frac{\lambda^2}{2}((b-a)^2/4)},
\]

where in the first step we have used the inequality we have seen in class for the Bernoulli random variable with parameter \( p \).

An alternative way to solve this problem could be define \( \phi(\lambda) = \ln \mathbb{E}[e^{\lambda X}] \).

\[
\phi'(\lambda) = \frac{d}{d\lambda} \ln \mathbb{E}[e^{\lambda X}] = \frac{\mathbb{E}[X e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]},
\]

So \( \phi(0) = \frac{0}{1} = 0 \).

\[
\phi''(\lambda) = \frac{d}{d\lambda} \phi'(\lambda) = \frac{d}{d\lambda} \frac{\mathbb{E}[X e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} = \frac{\mathbb{E}[X^2 e^{\lambda X} \mathbb{E}[e^{\lambda X}] - \mathbb{E}[X e^{\lambda X}]^2 \mathbb{E}[e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]^2}.
\]

6
For $\lambda = 0$, we have
\[ \phi''(0) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \text{Var}(X) \]
Also, we have $\phi(\lambda) \leq \phi(0) + \phi'(0)\lambda + \phi''(0)\frac{\lambda^2}{2} = \frac{\lambda^2}{2} \text{Var}(X)$ As $X$ is random variable taking values in $[a, b]$. The largest variance is achieved when $\Pr\{X = a\} = \frac{b}{b-a}$ $\Pr\{X = b\} = \frac{a}{b-a}$.

\[ \text{Var}(X) \leq \frac{(b-a)^2}{4} \]  

(16)
Therefore we have
\[ \mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2 (b-a)^2}{4}} \]

$X$ is $[(b-a)^2/4]$-subgaussian.

**Problem 5: Expected Maximum of Subgaussians**

Let $\{X_i\}_{i=1}^n$ be a collection of $n$ $\sigma^2$-subgaussian random variables, not necessarily independent of each other. Let $Y = \max_{i \in \{1, 2, \ldots, n\}} X_i$. Prove that $\mathbb{E}[Y] \leq \sqrt{2\sigma^2 \log n}$. Hint: Recall that by Jensen, $e^{\lambda \mathbb{E}[X]} \leq \mathbb{E}[e^{\lambda X}]$.

**Solution 5.** Consider the MGF of $Y$, we have the following relations for all $\lambda \geq 0$
\[ \mathbb{E}[e^{\lambda Y}] = \mathbb{E}[\exp(\lambda \max_{i \in \{1, 2, \ldots, n\}} X_i)] \leq \mathbb{E}\left[ \sum_{i \in \{1, 2, \ldots, n\}} e^{\lambda X_i} \right]. \]

Note that by the linearity of expectation (this does not require independence) and the assumptions that $\{X_i\}_{i=1}^n$ are $\sigma^2$-subgaussian random variables, we have
\[ \mathbb{E}[e^{\lambda Y}] \leq ne^{\lambda^2 \sigma^2/2}. \]

Using the hints, we have
\[ e^{\lambda \mathbb{E}[Y]} \leq e^{\lambda^2 \sigma^2/2 + \log n}, \]
which implies that
\[ \mathbb{E}[Y] \leq \frac{\sigma^2}{2} + \frac{1}{\lambda} \log n. \]

Optimizing over $\lambda$, we have the optimal $\lambda^* = \sqrt{\frac{2 \log n}{\sigma^2}}$, which gives us the desired inequality.