EE5137 2021/22 (Sem 2):
Solutions to Quiz 1 (Total 25 points)

Name:______________________________

Matriculation Number:_______________

Score:______________________________

You have 1.0 hours for this quiz. There are SIX (6) printed pages. You're allowed 1 sheet of handwritten notes. Please provide *careful explanations* for all your solutions.

1. Let X and Y be two independent Bernoulli (i.e., $\{0,1\}$-valued) random variables with

$$\Pr(X = 1) = \Pr(Y = 1) = 1/2.$$

   (a) (2 points) Are the random variables $X + Y \in \{0,1,2\}$ and $|X - Y| \in \{0,1\}$ independent? Explain carefully.

(b) (3 points) We say that two random variables A and B are *uncorrelated* if

$$\mathbb{E}[AB] = \mathbb{E}[A]\mathbb{E}[B].$$

Are the random variables $X+Y$ and $|X-Y|$ uncorrelated? Explain carefully.

Solution:

(a) No. They are not independent. If we know that $X+Y=2$, then both X and Y are equal to 1 and so $|X-Y|=0$. Thus, knowledge of $X+Y$ gives us information about $|X-Y|$.

(b) Let $A = X+Y$ and $B = |X-Y|$. We claim that they are uncorrelated. Note that $\mathbb{E}X = \mathbb{E}Y = 1/2$ and so $\mathbb{E}A = 1$. Note that $B=1$ with probability $1/2$ so that $\mathbb{E}B = 1/2$. Thus $(\mathbb{E}A)(\mathbb{E}B) = 1/2$. We want to check that $\mathbb{E}[AB] = \mathbb{E}[(X+Y)B] = 1/2$. We check $\mathbb{E}[XB]$. The joint distribution of X and $B = |X-Y|$ is

$$p_{X,B}(x,b) = \begin{cases} 1/4 & x=0, b=0 \\ 1/4 & x=0, b=1 \\ 1/4 & x=1, b=0 \\ 1/4 & x=1, b=1 \end{cases}$$

so

$$\mathbb{E}[XB] = \sum_{x,b} x\, b\, p_{X,B}(x,b) = 1/4$$

Thus, $\mathbb{E}[AB] = \mathbb{E}[XB] + \mathbb{E}[YB] = 1/2$. Since $\mathbb{E}[AB] = (\mathbb{E}A)(\mathbb{E}B) = 1/2$, we conclude that A and B are uncorrelated.

2. In machine learning and statistics, sub-Gaussian random variables play very important roles. We say that a zero-mean random variable X is *sub-Gaussian with variance proxy* σ^2, written as $X \sim \mathrm{subG}(\sigma^2)$, if its moment generating function $g_X(r)$ satisfies

$$g_X(r) = \mathbb{E}[e^{rX}] \leq \exp\left(\frac{r^2\sigma^2}{2}\right) \qquad \forall r \in \mathbb{R}.$$

(a) (2 points) If $X_i \sim \mathrm{subG}(\sigma_i^2)$ and the X_i's are zero-mean and independent, then what is the (smallest) variance proxy of $\sum_{i=1}^n X_i$?

(b) (3 points) If $X \sim \mathrm{subG}(\sigma^2)$ with zero mean, show that for any $t \geq 0$,

$$\Pr(X \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

(c) (5 points) Let $X \sim \mathrm{subG}(\sigma^2)$ with zero mean. Fix an integer $k \geq 1$. Use part (b) to find the best functions $f(k, \sigma^2)$ and $g(k)$ (i.e., those resulting in the tightest bound) such that

$$\mathbb{E}\left[|X|^k\right] \leq f(k, \sigma^2)\, \Gamma\big(g(k)\big) \quad \text{where} \quad \Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, \mathrm{d}x.$$

Solution:

(a) We compute the MGF of $\sum_{i=1}^n X_i$. We have

$$\mathbb{E}[e^{r\sum_{i=1}^n X_i}] = \prod_{i=1}^n \mathbb{E}[e^{rX_i}] = \prod_{i=1}^n \exp\left(\frac{r^2\sigma_i^2}{2}\right) = \exp\left(\frac{r^2}{2}\sum_{i=1}^n \sigma_i^2\right).$$

Hence, the variance proxy of $\sum_{i=1}^n X_i$ is $\sum_{i=1}^n \sigma_i^2$.

(b) Consider for fixed $r \geq 0$,

$$\Pr(X \geq t) = \Pr\left(e^{rX} \geq e^{rt}\right) \leq \frac{\mathbb{E}[e^{rX}]}{e^{rt}} \leq \exp(-rt)\exp\left(\frac{r^2\sigma^2}{2}\right).$$

Minimizing $r \mapsto -rt + r^2\sigma^2/2$ over all $r \geq 0$, we get $r^* = t/\sigma > 0$ (because $t, \sigma > 0$) so the exponent is $-r^*t + (r^*)^2\sigma^2/2 = -t^2/(2\sigma^2)$, i.e.,

$$\Pr(X \geq t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

as desired.

(c) Fix $k \in \mathbb{N}$. Note that if $X \sim \mathrm{subG}(\sigma^2)$, it also holds that $-X \sim \mathrm{subG}(\sigma^2)$. Then by the layer-cake representation,

$$\begin{aligned}\mathbb{E}\big[|X|^k\big] &= \int_0^\infty \Pr(|X|^k \geq t)\,\mathrm{d}t = \int_0^\infty \Pr(|X| \geq t^{1/k})\,\mathrm{d}t \\ &\leq 2\int_0^\infty \Pr(X \geq t^{1/k})\,\mathrm{d}t \leq 2\int_0^\infty \exp\left(-\frac{t^{2/k}}{2\sigma^2}\right)\,\mathrm{d}t.\end{aligned}$$

Now take $u = t^{2/k}/(2\sigma^2)$. Then we have

$$\mathbb{E}\big[|X|^k\big] \leq (2\sigma^2)^{k/2} k \int_0^\infty e^{-u} u^{k/2-1}\,\mathrm{d}u = (2\sigma^2)^{k/2}\, k\, \Gamma\left(\frac{k}{2}\right).$$

Hence, $f(k, \sigma^2) = (2\sigma^2)^{k/2}\, k$ and $g(k) = k/2$.

3. Let Y be a uniform random variable on $[0, 1]$. Given Y, we then toss a coin with bias Y repeatedly (i.e., the probability of seeing Head equals Y). The outcomes of the coin tosses are denoted by $X_1, X_2, \ldots \in \{H, T\}$.

   (a) (5 points) Suppose that among the first 2 coin tosses, one is Head and one is Tail. Find the conditional cumulative distribution function of Y, i.e., find

   $$F_{Y|\{X_1,X_2\}=\{H,T\}}(y) := \Pr\left(Y \leq y \mid \{X_1, X_2\} = \{H, T\}\right) \qquad \forall\, y \in [0, 1].$$

   Hint: Figuring out $\Pr(\{X_1, X_2\} = \{H, T\})$ *first would get you some marks. Think of using iterated expectations.*

(b) (5 points) Suppose that among the first n coin tosses, we observe k Heads. What is the probability that the $(n+1)$-st coin toss shows Head? More precisely, compute

$$\Pr\left(X_{n+1} = H \mid k \ \text{ Heads among } X_1, \ldots, X_n\right).$$

Hint: You can assume the following without proof. For $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$,

$$\int_0^1 y^k (1-y)^{n-k} \, \mathrm{d}y = \frac{k!(n-k)!}{(n+1)!}.$$

Solution:

(a) First, we have

$$\begin{aligned}
\Pr\left(\{X_1, X_2\} = \{H, T\}\right) &= \Pr\left(X_1 = H, X_2 = T\right) + \Pr\left(X_1 = T, X_2 = H\right) \\
&= 2\Pr\left(X_1 = H, X_2 = T\right) \\
&= 2\mathbb{E}\left[\Pr\left(X_1 = H, X_2 = T \mid Y\right)\right] \\
&= 2\mathbb{E}[Y(1-Y)] = 2\int_0^1 y(1-y)\,\mathrm{d}y = \frac{1}{3}.
\end{aligned}$$

Then, for $y \in [0,1]$ we have

$$\begin{aligned}
\Pr\left(\{X_1, X_2\} = \{H, T\}, Y \le y\right) &= 2\mathbb{E}\left[\mathbb{E}\left[1_{\{X_1 = H, X_2 = T\}} 1_{\{Y \le y\}} \mid Y\right]\right] \\
&= 2\mathbb{E}\left[1_{\{Y \le y\}} \mathbb{E}\left[1_{\{X_1 = H, X_2 = T\}} \mid Y\right]\right] \\
&= 2\mathbb{E}\left[1_{\{Y \le y\}} Y(1-Y)\right] \\
&= 2\int_0^y u(1-u)\,\mathrm{d}u = 2\left(\frac{y^2}{2} - \frac{y^3}{3}\right).
\end{aligned}$$

Therefore, for $y \in [0,1]$

$$\begin{aligned}
F_{Y \mid \{X_1, X_2\} = \{H, T\}}(y) &:= \Pr\left(Y \le y \mid \{X_1, X_2\} = \{H, T\}\right) \\
&= \frac{2\left(\frac{y^2}{2} - \frac{y^3}{3}\right)}{\frac{1}{3}} = 6\left(\frac{y^2}{2} - \frac{y^3}{3}\right).
\end{aligned}$$

(b) Similarly, we need to use the definition of conditional probability.
Look at the denominator first. Conditioned on Y, random variables $X_1, X_2, \ldots$

are i.i.d. Therefore, we have

$$
\begin{aligned}
\Pr\left(k \text{ Heads among } X_1, \ldots, X_n\right) &= \mathbb{E}\left[\Pr\left(k \text{ Heads among } X_1, \ldots, X_n \mid Y\right)\right] \\
&= \mathbb{E}\left[\binom{n}{k} Y^k (1-Y)^{n-k}\right] \\
&= \binom{n}{k} \int_0^1 y^k (1-y)^{n-k} \, \mathrm{d}y \\
&= \frac{n!}{k!(n-k)!} \frac{k!(n-k)!}{(n+1)!} \\
&= \frac{1}{n+1}
\end{aligned}
$$

Similarly, we obtain the numerator

$$
\begin{aligned}
\Pr\left(X_{n+1} = H,\ k \text{ Heads among } X_1, \ldots, X_n\right) &= \mathbb{E}\left[\binom{n}{k} Y^{k+1} (1-Y)^{n-k}\right] \\
&= \binom{n}{k} \int_0^1 y^{k+1} (1-y)^{n-k} \, \mathrm{d}y \\
&= \frac{n!}{k!(n-k)!} \frac{(k+1)![(n+1)-(k+1)]!}{[(n+1)+1]!} \\
&= \frac{k+1}{(n+1)(n+2)}
\end{aligned}
$$

The trick of how to utilize the formula provided is that k is replaced by $k+1$ and n replaced by $n+1$.

Therefore,

$$
\begin{aligned}
&\Pr\left(X_{n+1} = H \mid k \text{ Heads among } X_1, \ldots, X_n\right) \\
&= \frac{\Pr\left(X_{n+1} = H,\ k \text{ Heads among } X_1, \ldots, X_n\right)}{\Pr\left(k \text{ Heads among } X_1, \ldots, X_n\right)} = \frac{k+1}{n+2}.
\end{aligned}
$$