SI140 Probability & Statistics (Fall 2021): Final Exam Solutions

Problem 1

(10 points) Basic Concept.

(a) (5 points) Please describe the difference and connection between probability and statistics.

(b) (5 points) Please describe the pros and cons of Bayesian statistical inference and classical statistical inference.

Solution

(a) *Difference*: Probability deals with the prediction of future events. On the other hand, statistics are used to analyze the frequency of past events. More specifically, probability is a measure of the likelihood of an event to occur. In probability theory we consider some underlying process which has some randomness or uncertainty modeled by random variables, and we figure out what happens. In statistics we observe something that has happened, and try to figure out what underlying process would explain those observations.

We can also use the example of a jar of red and green jelly beans to show the difference between probability and statistics: A probabilist starts by knowing the proportion of each and asks the probability of drawing a red jelly bean. A statistician infers the proportion of red jelly beans by sampling from the jar.

Connection: Probability is used to answer questions in the category of Statistics. Most statistical models are based on experiments and hypotheses, and probability is integrated into the theory, to explain the scenarios better.

(b) *Pros of Bayesian statistical inference*:

- It provides a natural and principled way of combining prior information with data, within a solid decision theoretical framework. You can incorporate past information about a parameter and form a prior distribution for future analysis. When new observations become available, the previous posterior distribution can be used as a prior. All inferences logically follow from Bayes' theorem.
- It can calculate the probability that a hypothesis is true, which is generally what the researchers actually want to know.

Cons of Bayesian statistical inference:

- It requires you to know or construct a prior, but it does not tell you how to select the prior. If you do not proceed with caution, you can generate misleading results.
- It can produce posterior distributions that are heavily influenced by the priors.
- It may be computationally intensive due to integration over many parameters.

Pros of classical statistical inference:

- It tends to be less computationally intensive than the Bayesian statistical inference.

Cons of classical statistical inference:

- It does not take into account priors.
- It has only one well-defined hypothesis.

Problem 2

(10 points) Let X be a discrete r.v. whose distinct possible values are $x_0, x_1, \dots,$ and let $p_k = P(X = x_k)$. The entropy of X is $H(X) = \sum_{k=0}^{\infty} p_k \log_2(1/p_k)$.

(a) (5 points) Find $H(X)$ for $X \sim \text{Geom}(p)$.

(b) (5 points) Let X and Y be i.i.d. discrete r.v.s. Show that $P(X = Y) \geq 2^{-H(X)}$.

Solution

(a) By defining $q = 1 - p$, we have

$$H(X) = -\sum_{k=0}^{\infty}(pq^k)\log_2(pq^k) = -\log_2(p)\sum_{k=0}^{\infty} pq^k - \log_2(q)\sum_{k=0}^{\infty} kpq^k = -\log_2(p) - \frac{q}{p}\log_2(q).$$

(b) Let W be an r.v. taking value p_k with probability p_k. By Jensen, $\text{E}(\log_2(W)) \leq \log_2(\text{E}(W))$. But

$$\text{E}(\log_2(W)) = \sum_k p_k \log_2(p_k) = -H(X),$$
$$\text{E}W = \sum_k p_k^2 = P(X = Y),$$

so $-H(X) \leq \log_2 P(X = Y)$. Thus, $P(X = Y) \geq 2^{-H(X)}$.

Problem 3

(10 points) Let $X_1 \sim \text{Expo}(\lambda_1)$, $X_2 \sim \text{Expo}(\lambda_2)$ and $X_3 \sim \text{Expo}(\lambda_3)$ be independent.

(a) (5 points) Find $\text{E}(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3)$ in terms of $\lambda_1, \lambda_2, \lambda_3$.

(b) (5 points) Find $P(X_1 = \min(X_1, X_2, X_3))$.

Solution

(a) Since X_1, X_2 and X_3 are independent, we have

$$\begin{aligned}
&\text{E}\left(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3\right) \\
&= \text{E}\left(X_1 | X_1 > 1, X_2 > 2, X_3 > 3\right) + \text{E}\left(X_2 | X_1 > 1, X_2 > 2, X_3 > 3\right) + \text{E}\left(X_3 | X_1 > 1, X_2 > 2, X_3 > 3\right) \\
&= \text{E}\left(X_1 | X_1 > 1\right) + \text{E}\left(X_2 | X_2 > 2\right) + \text{E}\left(X_3 | X_3 > 3\right).
\end{aligned}$$

According to the memoryless property of exponential distribution, for each $i \in \{1, 2, 3\}$, we have

$$\text{E}\left(X_i | X_i > i\right) = \text{E}\left(X_i\right) + i = \frac{1}{\lambda_i} + i.$$

Then it follows that

$$\text{E}\left(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3\right) = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + 6.$$

(b) Note that we have

$$P\left(X_1 = \min\left(X_1, X_2, X_3\right)\right) = P\left(X_1 \leq \min\left(X_2, X_3\right)\right).$$

Since $\min\left(X_2, X_3\right) \sim \text{Expo}\left(\lambda_2 + \lambda_3\right)$, then by the property of exponential distribution, we have

$$P\left(X_1 = \min\left(X_1, X_2, X_3\right)\right) = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}.$$

Problem 4

(15 points) Let $X \sim \text{Gamma}(a, \lambda)$, $Y \sim \text{Gamma}(b, \lambda)$. Assume X and Y are independent.

(a) (5 points) Find the joint distribution of $T = X + Y$ and $W = \frac{X}{X+Y}$.

(b) (5 points) Find the distribution of T and W respectively.

(c) (5 points) Find $\text{E}(W)$.

Solution

(a) First, we calculate the joint PDF of T and W via the change of variables in two dimensions, *i.e.*, $f_{T,W}(t,w) = f_{X,Y}(x,y)\left|\frac{\partial(x,y)}{\partial(t,w)}\right|$. Let $t = x + y$ and $w = \frac{x}{x+y}$. Then we have $x = tw$ and $y = t(1-w)$. The Jacobian is

$$\frac{\partial(x,y)}{\partial(t,w)} = \begin{pmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial w} \end{pmatrix} = \begin{pmatrix} w & t \\ 1-w & -t \end{pmatrix}.$$

The absolution value of the determinant of the Jacobian is

$$\left|\begin{pmatrix} w & t \\ 1-w & -t \end{pmatrix}\right| = |-t| = t.$$

Consequently, $f_{T,W}(t,w) = f_{X,Y}(x,y) \cdot t = f_X(x) \cdot f_Y(y) \cdot t$, where the second equality comes from the independence of X and Y. Remind that $X \sim \text{Gamma}(a,\lambda)$, $Y \sim \text{Gamma}(b,\lambda)$ and $x = tw$, $y = t(1-w)$, thus we have

$$\begin{aligned}
&f_{T,W}(t,w) \\
=&f_X(x) \cdot f_Y(y) \cdot t \\
=&\frac{1}{\Gamma(a)}(\lambda x)^a e^{-\lambda x}\frac{1}{x} \cdot \frac{1}{\Gamma(b)}(\lambda y)^b e^{-\lambda y}\frac{1}{y} \cdot t \\
=&\frac{1}{\Gamma(a)}(\lambda t w)^a e^{-\lambda t w}\frac{1}{tw} \cdot \frac{1}{\Gamma(b)}(\lambda t(1-w)))^b e^{-\lambda t(1-w)}\frac{1}{t(1-w)} \cdot t \\
=&\frac{1}{\Gamma(a)\Gamma(b)}w^{a-1} \cdot (1-w)^{b-1} \cdot (\lambda t)^{a+b} e^{-\lambda t} \cdot \frac{1}{t} \\
=&\frac{1}{\Gamma(a+b)}(\lambda t)^{a+b} e^{-\lambda t}\frac{1}{t} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}w^{a-1}(1-w)^{b-1}.
\end{aligned}$$

(b) From the result of (a), we can find that the joint PDF of T and W can be written as the product of two functions that only depend on t and w, respectively:

$$g(t) = \frac{1}{\Gamma(a+b)}(\lambda t)^{a+b} e^{-\lambda t}\frac{1}{t}$$

and

$$h(w) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}w^{a-1}(1-w)^{b-1}.$$

We see that $g(t)$ and $h(w)$ are actually the probability density functions of $\text{Gamma}(a+b,\lambda)$ and $\text{Beta}(a,b)$, respectively. It follows that $T \sim \text{Gamma}(a+b,\lambda)$ and $W \sim \text{Beta}(a,b)$ according to the story of post-bank story.

(c) Since $W \sim \text{Beta}(a,b)$, we have $E(W) = \frac{a}{a+b}$.

Problem 5

(10 points) Instead of predicting a single value for the parameter, we give an interval that is likely to contain the parameter: A $1-\delta$ confidence interval for a parameter p is an interval $[\hat{p}-\epsilon, \hat{p}+\epsilon]$ such that $\Pr(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]) \geq 1-\delta$. Now we toss a coin with probability p landing heads and probability $1-p$ landing tails. The parameter p is unknown and we need to estimate its value from experiments results. We toss such coin N times. Let $X_i = 1$ if the ith result is head, otherwise 0. We estimate p by using $\hat{p} = \frac{X_1+\ldots+X_N}{N}$. Find the $1-\delta$ confidence interval for p, then discuss the impacts of δ and N. **Hint**: You can use the following Hoeffding bound: Let the random variables $X_1, X_2, \ldots, X_n$ be independent with $\mathrm{E}(X_i) = \mu$, $a \leq X_i \leq b$ for each $i = 1, \ldots, n$, where a, b are constants. Then for any $\epsilon \geq 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq 2e^{-\frac{2n\epsilon^2}{(b-a)^2}}.$$

Solution

Since $X_i \sim \mathrm{Bern}(p)$, then $\mathrm{E}(X_i) = p$. Furthermore, there are only two possible values of X_i, 0 and 1, so $0 \leq X_i \leq 1$. Then by Hoeffding bound, we have

$$\Pr(|\hat{p}-p| \geq \epsilon) = \Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N} X_i - p\right| \geq \epsilon\right) \leq 2e^{-2N\epsilon^2}.$$

Let $2e^{-2N\epsilon^2} = \delta$, we can get

$$\epsilon = \sqrt{\frac{\ln(\frac{2}{\delta})}{2N}}.$$

It follows that

$$\begin{aligned} &\Pr(|\hat{p}-p| \geq \epsilon) \leq \delta \\ \iff &\Pr(|\hat{p}-p| < \epsilon) > 1-\delta \\ \iff &\Pr(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]) < \epsilon) > 1-\delta \end{aligned}$$

where $\hat{p} = \frac{X_1+\ldots+X_N}{N}$, $\epsilon = \sqrt{\frac{\ln(\frac{2}{\delta})}{2N}}$.

Then we have found the confidence interval for p. It means that $\forall \delta > 0$, $|\hat{p}-p| < \sqrt{\frac{\ln(\frac{2}{\delta})}{2N}}$ with probability greater than $1-\delta$. We can see that if N rises, δ rises, then ϵ will drop, which means we can get a more accurate estimation for p.

Problem 6

(15 points) Given a coin with the probability p of landing heads. p is unknown and we need to estimate its value through data. In our data collection model, we have n independent tosses, result of each toss is either Head or Tail. Let X denote the number of heads in the total n tosses. Now we conduct experiments to collect data and find $X = k$. Then we need to find $\hat{p}$, the estimation of p.

(a) (5 points) Assume p is an unknown constant. Find $\hat{p}$ through the MLE (Maximum Likelihood Estimation) rule.

(b) (5 points) Assume p is a random variable with a prior distribution $p \sim \text{Beta}(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MAP (Maximum a Posterior Probability) rule.

(c) (5 points) Assume p is a random variable with a prior distribution $p \sim \text{Beta}(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MMSE (Minimal Mean Squared Error) rule.

Solution

(a) We know $X \sim \text{Bin}(n, p)$, where p is an unknown constant. The likehood function is

$$P_X(k; p) = \Pr(X = k|p) = \binom{n}{k} p^k (1-p)^{n-k}.$$

The corresponding log-likelihood function is

$$h(p) = \log(P_X(k; p)) = \binom{n}{k} [k \log p + (n-k) \log(1-p)].$$

Our goal is to find p^* such that $h(p^*)$ is the maximum of $h(p)$. We have

$$h'(p) = \binom{n}{k} \left[\frac{k}{p} - \frac{n-k}{1-p} \right],$$
$$h''(p) = \binom{n}{k} \left[-\frac{k}{p^2} - \frac{n-k}{(1-p)^2} \right] \le 0.$$

Let $h'(p) = 0$ we have $p^* = \frac{k}{n}$. $h(p^*)$ is the maximum of $h(p)$ since $h''(p) \le 0$. Therefore, by the MLE rule, $\hat{p}_{MLE} = \frac{k}{n}$.

(b) We know the posterior distribution

$$f_{p|X=k} \propto p^{a+k-1}(1-p)^{b+n-k-1},\ p \in (0, 1)$$

by Beta-Binomial conjugacy. Then the MAP estimator

$$\hat{p}_{MAP} = \arg\max_p f_{p|X=k} = \arg\max_p \log(f_{p|X=k})$$

since logarithmic function is monotonically increasing.
Let

$$g(p) = \log(f_{p|X=k}) = (a + k - 1) \log p + (b + n - k - 1) \log(1 - p),$$

where we don't consider the proportional constant. Our goal is to find p^* such that $g(p^*)$ is maximum of $g(p)$. We have

$$g'(p) = \frac{a+k-1}{p} - \frac{b+n-k-1}{1-p},$$
$$g''(p) = -\frac{a+k+1}{p^2} - \frac{b+n-k-1}{(1-p)^2} < 0.$$

Let $g'(p^*) = 0$. We have $p^* = \frac{a+k-1}{a+b+n-2}$, and $g(p^*)$ is maximum of $g(p)$ since $g''(p) < 0$.

Then we can get the MAP estimate

$$\hat{p}_{MAP} = \arg\max_p f_{p|X=k} = \arg\max_p \log(f_{\theta|X=k}) = p^* = \frac{a+k-1}{a+b+n-2}.$$

(c) Since the prior distribution is $p \sim \text{Beta}(a, b)$ and the conditional distribution of X given p is $X|p \sim \text{Bin}(n, p)$, we can get the posterior distribution

$$p|X = k \sim \text{Beta}(a + k, b + n - k)$$

by Beta-Binomial conjugacy. It follows that

$$\text{E}(p|X = k) = \frac{a+k}{a+b+n},$$

so the MMSE estimator of p is

$$\hat{p}_{MMSE} = E(p|X = k) = \frac{a+k}{a+b+n}.$$

Problem 7

(10 points) We know that the MMSE of X given Y is given by $g(Y) = \mathrm{E}[X|Y]$. We also know that the Linear Least Square Estimate (LLSE) of X given Y, denoted by $L[Y|X]$, is shown as follows:

$$L[Y|X] = \mathrm{E}(Y) + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}(X - \mathrm{E}(X))$$

Now we wish to estimate the probability of landing heads, denoted by θ, of a biased coin. We model θ as the value of a random variable Θ with a known prior PDF $f_\Theta \sim \mathrm{Unif}(0,1)$. We consider n independent tosses and let X be the number of heads observed.

(a) (5 points) Show that $\mathrm{E}[(\Theta - \mathrm{E}[\Theta|X])h(X)] = 0$ for any real function $h(\cdot)$.

(b) (5 points) Find the MMSE $\mathrm{E}[\Theta|X]$ and LLSE $L[\Theta|X]$. (Eve's law: $\mathrm{Var}(Y) = \mathrm{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathrm{E}(Y|X))$.)

Solution

(a) We can see that

$$\begin{aligned} &\mathrm{E}\left[\left(\Theta - \mathrm{E}\left[\Theta|X\right]\right)h\left(X\right)\right] = \mathrm{E}\left[\Theta h\left(X\right) - \mathrm{E}\left[\Theta|X\right]h\left(X\right)\right] \\ &= \mathrm{E}\left[\Theta h\left(X\right)\right] - \mathrm{E}\left[\mathrm{E}\left[\Theta|X\right]h\left(X\right)\right] = \mathrm{E}\left[\Theta h\left(X\right)\right] - \mathrm{E}\left[\mathrm{E}\left[\Theta h\left(X\right)|X\right]\right]. \end{aligned}$$

By Adam's law, we have $\mathrm{E}\left[\mathrm{E}\left[\Theta h\left(X\right)|X\right]\right] = \mathrm{E}\left[\Theta h\left(X\right)\right]$. Then it follows that

$$\mathrm{E}\left[\left(\Theta - \mathrm{E}\left[\Theta|X\right]\right)h\left(X\right)\right] = \mathrm{E}\left[\Theta h\left(X\right)\right] - E\left[\Theta h\left(X\right)\right] = 0.$$

(b) Firstly, we find the MMSE $\mathrm{E}[\Theta|X]$. Since the prior distribution

$$\Theta \sim \mathrm{Unif}(0,1) = \mathrm{Beta}(1,1)$$

and given $\Theta = \theta$, the conditional distribution of X

$$X|\Theta = \theta \sim \mathrm{Bin}(n,\theta),$$

by Beta-Binomial conjugacy we can get the posterior distribution

$$\Theta|X = k \sim \mathrm{Beta}(k+1, n-k+1).$$

It follows that

$$\mathrm{E}(\Theta|X = k) = \frac{k+1}{(k+1)+(n-k+1)} = \frac{k+1}{n+2},$$

so the MMSE estimator of Θ is

$$\mathrm{E}(\Theta|X) = \frac{X+1}{n+2}.$$

Then we find the LLSE $L[\Theta|X]$. Since

$$L[\Theta|X] = \mathrm{E}(\Theta) + \frac{\mathrm{Cov}(\Theta,X)}{\mathrm{Var}(X)}(X - \mathrm{E}(X)),$$

we just need to calculate these statistics. By $\Theta \sim \mathrm{Unif}(0,1)$ we have

$$\mathrm{E}(\Theta) = \frac{1}{2}, \quad \mathrm{Var}(\Theta) = \frac{1}{12}, \quad \mathrm{E}(\Theta^2) = \frac{1}{3}.$$

We know $X|\Theta = \theta \sim \mathrm{Bin}(n,\theta)$, so

$$\mathrm{E}(X|\Theta = \theta) = n\theta \implies \mathrm{E}(X|\Theta) = n\Theta,$$

$$\operatorname{Var}(X|\Theta=\theta) = n\theta(1-\theta) \implies \operatorname{Var}(X|\Theta) = n\Theta(1-\Theta).$$

By Adam's law, we have

$$\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X|\Theta)] = \mathrm{E}(n\Theta) = n\cdot\mathrm{E}(\Theta) = \frac{n}{2}.$$

By Eve's law, we have

$$\begin{aligned}
\operatorname{Var}(X) &= \mathrm{E}(\operatorname{Var}(X|\Theta)) + \operatorname{Var}(\mathrm{E}(X|\Theta)) \\
&= \mathrm{E}(n\Theta(1-\Theta)) + \operatorname{Var}(n\Theta) \\
&= n(\mathrm{E}(\Theta) - \mathrm{E}(\Theta^2)) + n^2\operatorname{Var}(\Theta) \\
&= n\left(\frac{1}{2} - \frac{1}{3}\right) + n^2\cdot\frac{1}{12} \\
&= \frac{n}{12}(n+2).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\operatorname{Cov}(X,\Theta) &= \mathrm{E}(\Theta X) - \mathrm{E}(\Theta)\mathrm{E}(X) \\
&= \mathrm{E}[\mathrm{E}(\Theta X|\Theta)] - \mathrm{E}(\Theta)\mathrm{E}(X) \\
&= \mathrm{E}(\Theta\mathrm{E}(X|\Theta)) - \mathrm{E}(\Theta)\mathrm{E}(X) \\
&= \mathrm{E}(\Theta\cdot n\Theta) - \mathrm{E}(\Theta)\mathrm{E}(X) \\
&= n\mathrm{E}(\Theta^2) - \mathrm{E}(\Theta)\mathrm{E}(X) \\
&= n\cdot\frac{1}{3} - \frac{1}{2}\cdot\frac{n}{2} \\
&= \frac{n}{12}.
\end{aligned}$$

It follows that

$$\begin{aligned}
L[\Theta|X] &= \mathrm{E}(\Theta) + \frac{\operatorname{Cov}(\Theta,X)}{\operatorname{Var}(X)}(X-\mathrm{E}(X)) \\
&= \frac{1}{2} + \frac{\frac{n}{12}}{\frac{n}{12}(n+2)}\left(X - \frac{n}{2}\right) \\
&= \frac{1}{2} + \frac{1}{n+2}\left(X - \frac{n}{2}\right) \\
&= \frac{X+1}{n+2}.
\end{aligned}$$

We can find in this case $\mathrm{E}[\Theta|X] = L[\Theta|X]$.

Problem 8

(10 points) Show the following inequalities.

(a) (5 points) Let $X \sim \text{Pois}(\lambda)$. If there exists a constant $a > \lambda$, then

$$\mathbb{P}(X \geq a) \leq \frac{e^{-\lambda}(e\lambda)^a}{a^a}.$$

(b) (5 points) Let X be a random variable with finite variance σ^2. Then for any constant $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{2\sigma^2}{\sigma^2 + a^2}.$$

Solution

(a) Using Chernoff's technique and Markov's inequality, we have $\forall t > 0$,

$$P(X \geq a) = P\left(e^{tX} \geq e^{ta}\right) \leq \frac{\mathrm{E}\left[e^{tX}\right]}{e^{ta}}.$$

By $X \sim \text{Pois}(\lambda)$, we have

$$\mathrm{E}\left[e^{tX}\right] = \sum_{k=0}^{\infty} e^{tk} \cdot e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda\left(e^t - 1\right)}.$$

Thus

$$P(X \geq a) \leq e^{\lambda\left(e^t-1\right)-ta},\ \forall t > 0.$$

When $t = \ln\frac{a}{\lambda}$, we can get $\min_{t>0} e^{\lambda\left(e^t-1\right)-ta} = \frac{e^{-\lambda}(e\lambda)^a}{a^a}$, thus

$$P(X \geq a) \leq \frac{e^{-\lambda}(e\lambda)^a}{a^a}.$$

(b) When $\sigma \geq a$, we have $\frac{2\sigma^2}{\sigma^2+a^2} \geq 1$, and it follows that

$$P(|X - \mathrm{E}[X]| \geq a) \leq 1 \leq \frac{2\sigma^2}{\sigma^2 + a^2}.$$

When $\sigma < a$, using Markov's inequality, we have

$$P(|X - \mathrm{E}[X]| \geq a) = P\left((X - \mathrm{E}[X])^2 \geq a^2\right) \leq \frac{\mathrm{E}\left[(X - \mathrm{E}[X])^2\right]}{a^2} = \frac{\sigma^2}{a^2}.$$

Since $\sigma < a$, we have $\frac{\sigma^2}{a^2} \leq \frac{2\sigma^2}{\sigma^2+a^2}$, thus

$$P(|X - \mathrm{E}[X]| \geq a) \leq \frac{2\sigma^2}{\sigma^2 + a^2}.$$

Problem 9

(10 points) Let $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$; X and Y are independent. Now let $Z_1 = \sin(X+Y)$, $Z_2 = \cos(X+Y)$.

(a) Find $\mathrm{E}(Z_1)$ and $\mathrm{E}(Z_2)$.

(b) Find $\mathrm{Var}(Z_1)$ and $\mathrm{Var}(Z_2)$.

Solution

(a) First, we find $\mathrm{E}(Z_1)$. Since $X+Y \sim \mathcal{N}(0,2)$ and $-(X+Y) \sim \mathcal{N}(0,2)$ have the same distribution, we have

$$\mathrm{E}\left(\sin\left(X+Y\right)\right) = \mathrm{E}\left(\sin\left(-\left(X+Y\right)\right)\right) = \mathrm{E}\left(-\sin\left(X+Y\right)\right) = -\mathrm{E}\left(\sin\left(X+Y\right)\right).$$

$$\Rightarrow \mathrm{E}\left(Z_1\right) = \mathrm{E}\left(\sin\left(X+Y\right)\right) = 0.$$

Next, we provide two different methods to find $E(Z_2)$.

Method 1: Since $X \sim \mathcal{N}(0,1)$, the MGF of the r.v. jX is

$$M_{jX}(t) = \mathrm{E}\left(e^{jXt}\right) = \int_{-\infty}^{\infty} e^{jxt} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = e^{-\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-jt)^2}{2}} dx = e^{-\frac{1}{2}t^2}.$$

Then we consider a r.v. $U \sim \mathcal{N}(0,\sigma^2)$. Since U has the same distribution as σX, we have

$$M_{jU}(t) = M_{j\sigma X}(t) = \mathrm{E}\left(e^{j\sigma X t}\right) = M_{jX}(\sigma t) = e^{-\frac{1}{2}(\sigma t)^2}.$$

When $t=1$, we obtain

$$M_{jU}(1) = \mathrm{E}\left(e^{jU}\right) = e^{-\frac{1}{2}\sigma^2}.$$

Besides, we have the following result

$$\mathrm{E}\left[e^{jU}\right] = \mathrm{E}\left[\cos U + j\sin U\right] = \mathrm{E}\left[\cos U\right] + j\mathrm{E}\left[\sin U\right].$$

Therefore, we conclude that for the r.v. $U \sim \mathcal{N}(0,\sigma^2)$,

$$\mathrm{E}\left[\cos U\right] = e^{-\frac{1}{2}\sigma^2},\ \mathrm{E}\left[\sin U\right] = 0.$$

Finally, by $X+Y \sim \mathcal{N}(0,2)$, we have

$$\mathrm{E}\left[Z_2\right] = \mathrm{E}\left(\cos\left(X+Y\right)\right) = e^{-1}.$$

Method 2: Consider a r.v. $U \sim \mathcal{N}\left(0,\sigma^2\right)$, its Taylor series expansion is

$$\cos U = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} U^{2n}.$$

Using GMF we obtain $\mathrm{E}[U^{2n}] = \frac{(2n)!\sigma^{2n}}{2^n n!} = (2n-1)!!\sigma^{2n}$, where $(2n-1)!! = 1 \cdot 3 \cdot 5 \cdots (2n-1)$. Then we have

$$\mathrm{E}\left(\cos U\right) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} \mathrm{E}\left(U^{2n}\right) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} \cdot \sigma^{2n}\left(2n-1\right)!! = \sum_{n=0}^{\infty} \frac{(-1)^n \sigma^{2n}}{2^n n!} = \sum_{n=0}^{\infty} \frac{\left(-\sigma^2/2\right)^n}{n!} = e^{-\sigma^2/2},$$

By $X+Y \sim \mathcal{N}(0,2)$, we have

$$\mathrm{E}\left(Z_2\right) = \mathrm{E}\left(\cos(X+Y)\right) = e^{-1}.$$

(b) Since $2(X+Y) \sim \mathcal{N}(0,8)$, we have

$$\mathrm{E}\left(Z_2^2\right) - \mathrm{E}\left(Z_1^2\right) = \mathrm{E}\left(Z_2^2 - Z_1^2\right) = \mathrm{E}\left(\cos^2(X+Y) - \sin^2(X+Y)\right) = \mathrm{E}\left(\cos\left(2\left(X+Y\right)\right)\right) = e^{-4}.$$

Besides, we have

$$\mathrm{E}\left(Z_2^2\right) + \mathrm{E}\left(Z_1^2\right) = \mathrm{E}\left(Z_2^2 + Z_1^2\right) = \mathrm{E}\left(\cos^2(X+Y) + \sin^2(X+Y)\right) = 1.$$

Then we can solve for $E(Z_1^2)$ and $E(Z_2^2)$ as

$$\mathrm{E}\left(Z_1^2\right) = \frac{1-e^{-4}}{2},\ \mathrm{E}\left(Z_2^2\right) = \frac{1+e^{-4}}{2}.$$

Finally, we obtain the following results

$$\begin{aligned}
\mathrm{Var}\left(Z_1\right) &= \mathrm{E}\left(Z_1^2\right) - \mathrm{E}^2\left(Z_1\right) = \frac{1-e^{-4}}{2}, \\
\mathrm{Var}\left(Z_2\right) &= \mathrm{E}\left(Z_2^2\right) - \mathrm{E}^2\left(Z_2\right) = \frac{1+e^{-4}}{2} - e^{-2} = \frac{\left(1-e^{-2}\right)^2}{2}.
\end{aligned}$$