

Problems

The expressions for the probability mass function of the Poisson(λ) distribution, and the density function of the Normal distribution with mean μ and variance σ^2 , may be useful:

$$\frac{\lambda^n}{n!} e^{-\lambda} \quad \text{and} \quad \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

1. (10 points) For each of the following questions, exactly one answer is correct. Each correct answer gives 1 point, and each incorrect answer results in a 1/2 point reduction. The minimal possible total score for the full problem is 0. *Some of the results may be useful in later problems!*
- a) Let X_1, X_2, \dots be an i.i.d. sequence of random variables with $0 < \mathbb{E}(X_1^2) < \infty$. Which of the following statements is a consequence of the strong law of large numbers?
1. $\limsup_{n \rightarrow \infty} X_n = +\infty$ almost surely.
 2. $\lim_{n \rightarrow \infty} (X_1^2 + \dots + X_n^2)/n = 0$ almost surely.
 3. $\sup_{n \geq 1} |X_n| < \infty$ almost surely.
- b) Let φ_1 and φ_2 be the characteristic functions of two random variables X_1 and X_2 . What does the identity theorem for characteristic functions tell us?
1. If $\varphi_1(u) = \varphi_2(u)$ for all $u \in \mathbb{R}$ then X_1 and X_2 have the same distribution.
 2. If $\varphi_1(u) = \varphi_2(u)$ for all $u \in \mathbb{R}$ then X_1 and X_2 are equal almost surely.
 3. If $\varphi_1(u) = \varphi_2(u)$ for all $u \in \mathbb{R}$ then X_1 and X_2 are independent.
- c) What does it mean for a statistical test to have significance level 0.05?
1. The Null hypothesis will be accepted with probability at least 0.05.
 2. If the Null hypothesis is true, then the probability to reject is at most 0.05.
 3. If the Null hypothesis is false, then the probability to reject is at least 0.95.
- d) The characteristic function φ of a Poisson(λ) random variable is given by
1. $\varphi(u) = \exp(u(e^{i\lambda} - 1))$.
 2. $\varphi(u) = \exp(\lambda(e^{iu} - 1))$.
 3. $\varphi(u) = \exp(\lambda(iu - 1))$.

- e) Suppose a statistical test resulted in a p-value of 0.07. Which of the following statements is correct?
1. The probability that the Null hypothesis is false is 0.07.
 2. The Null hypothesis could not be rejected at significance level 0.06.
 3. The Null hypothesis could not be rejected at significance level 0.08.
- f) Suppose X_1, \dots, X_n are i.i.d. $N(\mu, \sigma^2)$. What is the distribution of $\frac{1}{n} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$?
1. $N(0, 1)$.
 2. $N(0, 1/n)$.
 3. Student t with n degrees of freedom.
- g) Suppose X_1, \dots, X_n are i.i.d. Bernoulli(p) for some $0 < p < 1$. What is the distribution of $X_1 + \dots + X_n$?
1. Binomial(n, p).
 2. Poisson(np).
 3. Geometric(n, p).
- h) Which of the following formulas is **not** correct in general?
1. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
 2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap B)$.
 3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(A \cap B^c)$.
- i) Which of the following statements is correct for any random variables X, Y with unit variance, $\text{Var}(X) = \text{Var}(Y) = 1$?
1. If $\text{Cov}(X, Y) = 0$ then X and Y are independent.
 2. If $X = a + bY$ for some constants $a \in \mathbb{R}$ and $b > 0$, then $\text{Cov}(X, Y) = 1$.
 3. If $X = Y^2$, then $\text{Cov}(X, Y) = 1$.
- j) What is the value of $\sum_{k=0}^{n-1} 2^k$?
1. $2^n - 1$.
 2. $n^2 - 1$.
 3. $2^n - n$.

2. (10 points) A team of particle physicists has performed a series of experiments using the Large Hadron Collider (LHC) to study Higgs boson production. The total number of produced Higgs bosons is $N \sim \text{Poisson}(\lambda)$. Unfortunately, due to background effects, other particles are also produced during the experiments. The total number of such particles is $M \sim \text{Poisson}(\gamma)$, with M and N independent. The parameters λ and γ are nonnegative and unknown.

The detector unit can only measure the total number of produced particles, $X = M + N$.

- a) Show that $X \sim \text{Poisson}(\lambda + \gamma)$.

The parameter of interest is λ , which controls the number of Higgs bosons produced. To proceed, further information about the background parameter γ is needed. Therefore, the physicists perform a separate experiment from which a random variable $Y \sim N(\gamma, \sigma^2)$ is observed independently of X , where $\sigma^2 > 0$ is known.

- b) The joint distribution of (X, Y) given the parameters $\theta = (\lambda, \gamma)$ is of the form

$$\mathbb{P}_\theta(X = n, Y \leq z) = \int_{-\infty}^z f(n, y; \theta) dy.$$

Determine $f(n, y; \theta)$.

Consider the likelihood function $L(\theta) = f(X, Y; \theta)$.

- c) Find the maximum likelihood estimate $(\hat{\lambda}, \hat{\gamma})$ of (λ, γ) , knowing X and Y , in the case where $0 < Y < X$.

3. (15 points) Consider a series of coin tosses modeled by a sequence X_1, X_2, \dots of i.i.d. random variables with $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$. Here $X_n = 1$ signifies heads and $X_n = -1$ tails. Before each toss, you have the opportunity to bet an amount V_n . If the coin comes up heads ($X_n = 1$), you receive V_n . Otherwise you lose V_n . Therefore the total gain up to and including the n th toss is

$$G_n = \sum_{k=1}^n V_k X_k.$$

Note that G_n may be negative, which signifies a loss.

Suppose you use the following strategy: Set $V_1 = 1$. For $k \geq 1$, if $X_k = 1$ you stop, meaning that you set $V_n = 0$ for $n \geq k + 1$. If $X_k = -1$ and you have not yet stopped, double your bet, that is, set $V_{k+1} = 2V_k$.

Furthermore, let T denote the time you stop, $T = \min\{k \geq 1 : X_k = 1\}$, with $T = \infty$ if $X_k = -1$ for all k .

- a) Compute G_n for $1 \leq n < T$.
- b) Compute G_T for $T < \infty$.
- c) Compute $\mathbb{P}(T < \infty)$.

At this point you may feel that this is too good to be true—and indeed there is a catch:

- d) Compute the expected maximal intermediate shortfall. That is, compute $\mathbb{E}\left(\min_{1 \leq n \leq T} G_n\right)$.

To make matters worse, in reality you would face a limit on credit. Specifically, assume you are only allowed to use strategies $\{V_n : n \geq 1\}$ such that the gain satisfies $G_n \geq -b$ almost surely for all n , where $b \in (1, \infty)$ is some fixed constant. Modify the above strategy so that you stop as soon as there is a risk of violating the bound. That is, specify V_k as before, except that if you have not stopped at k , and $G_k - 2V_k < -b$, then set $V_n = 0$ for all $n \geq k + 1$.

Let T' denote the time you stop under this modified strategy.

- e) Compute your expected gain $\mathbb{E}(G_{T'})$ under these new rules. In particular, how does this depend on the bound b ?

4. (15 points) Celebrities Saylor Twift and Bustin Jieber are active on Twitter. Saylor has a total of n followers. There is some overlap between Twist and Jieber followers: a fraction $\gamma \in (0, 1)$ of Saylor’s followers also follow Bustin (that is, $[\gamma n]$ people follow both Saylor and Bustin, where $[x]$ denotes the integer part of a real number x).

Let A denote the event that Bustin tweets “Saylor Twift got a #beautifulvoice”. If a follower of Saylor reads this tweet, then this follower will re-tweet Saylor’s tweets with probability $p_S = 0.3$. Otherwise, the probability is only $q_S = 0.1$. Followers make tweeting decisions independently of each other, and do not re-tweet each other’s tweets.

One day, Saylor tweets “today glimpsed a new world who knew that #statsrulez?”. Let N denote the total number of re-tweets of Saylor’s tweet.

- a) Assuming that A **did not** happen, find $\mathbb{E}(N)$ and $\text{Var}(N)$.
- b) Assuming that A **did** happen, find $\mathbb{E}(N)$ and $\text{Var}(N)$.

Suppose you work for Twitter and a colleague tells you the value of N . You are currently offline and cannot check specific tweets. You are interested in if A happened or not.

- c) Under each of the distributions in a) and b), show that

$$\frac{N - \mathbb{E}(N)}{\sqrt{\text{Var}(N)}}$$

converges weakly to the standard Normal distribution as $n \rightarrow \infty$, with p_S , q_S , and γ held fixed.

Hint: It may be helpful to write N as a sum of Bernoulli random variables. Furthermore, you are allowed to use the following result:

Lemma: Let U_1, U_2, \dots and V_1, V_2, \dots be two sequences of random variables that both converge weakly to the standard Normal distribution. Suppose U_n and V_n are independent for each n . Let $\varrho_n \in (0, 1)$ and assume $\lim_n \varrho_n = \varrho \in (0, 1)$. Then $\varrho_n U_n + \sqrt{1 - \varrho_n^2} V_n$ converges weakly to the standard Normal distribution.

- d) Consider the hypotheses

$$\begin{cases} \text{Null:} & \text{The event } A \text{ happened;} \\ \text{Alternative:} & \text{The event } A \text{ did not happen.} \end{cases}$$

In this problem, you should approximate the corresponding distributions of N by Normals with the same mean and variance. Show that one can find a most powerful test at a given level α of the form

$$\text{reject the Null} \iff |N - 0.05n| > c$$

for some constant c . (You do not have to determine c ; it depends on α and the distribution of N under the Null hypothesis.)

5. (15 points) A common measure of the quality of an estimator is the mean squared error (MSE). In this problem, you will see that the MSE may behave in unexpected ways. This was first noticed by Stein in 1956 and developed further by James and Stein in 1961.

Fix $n \geq 2$. Let $X = (X_1, \dots, X_n)$ be a random vector, where X_1, \dots, X_n are independent with $X_i \sim N(\theta_i, 1)$ for some unknown parameters $\theta = (\theta_1, \dots, \theta_n)$.

- a) Consider the estimator $\hat{\theta}_i = X_i$ for $i = 1, \dots, n$, and set $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_n)$. Compute the mean squared error

$$\mathbb{E}(\|\theta - \hat{\theta}\|^2)$$

Consider now the alternative estimator $\hat{\theta}^{\text{JS}} = (\hat{\theta}_1^{\text{JS}}, \dots, \hat{\theta}_n^{\text{JS}})$, where

$$\hat{\theta}_i^{\text{JS}} = X_i - \frac{n-2}{\|X\|^2} X_i, \quad i = 1, \dots, n.$$

- b) Let $h : \mathbb{R}^n \rightarrow \mathbb{R}$ be continuously differentiable and such that $h(x) = 0$ whenever $\|x\|$ is sufficiently large (that is, for $h(x) = 0$ whenever $\|x\| > C$ for some constant C). Show that

$$\mathbb{E}\left((X_i - \theta_i)h(X)\right) = \mathbb{E}\left(\frac{\partial h}{\partial x_i}(X)\right) \quad \text{for all } i = 1, \dots, n. \quad (1)$$

Hint: Use that $\phi'(t) = -t\phi(t)$, where ϕ denotes the $N(0, 1)$ density function.

Equation (1) actually holds for more general functions h . In particular, for any $i \in \{1, \dots, n\}$, one can take $h(x) = x_i/|x|^2$ (setting $h(0) = 0$ and $\frac{\partial h}{\partial x_i}(0) = 0$), as long as $n \geq 3$. You may use this fact without proof.

- c) For $n \geq 3$, show that $\hat{\theta}^{\text{JS}}$ has smaller mean squared error than $\hat{\theta}$. That is, show that

$$\mathbb{E}(\|\theta - \hat{\theta}^{\text{JS}}\|^2) < \mathbb{E}(\|\theta - \hat{\theta}\|^2).$$

This result is counterintuitive, because the X_i are independent. It is tempting to conclude that “the quality of an estimate can be improved by simultaneously estimating independent variables”. Food for thought...