1. Suppose that 10 points, $P_1, P_2, \ldots, P_{10}$, are independently chosen at random on the perimeter of a circle. Let $S$ denote the event that all the points are contained in some semicircle, i.e., there is a line passing through the center of the circle such that all the points are on one side of that line. Let $S_i$ be the event that all the points lie in the semicircle beginning at the point $P_i$ and going clockwise for $180^\circ$, $i = 1, 2, \ldots, 10$. (a) Express $S$ in terms of the $S_i$. (b) Find $P(S)$.

解答:

(a) $S = \bigcup_{i=1}^{10} S_i$.

(b) Note that $S_1, S_2, \ldots, S_{10}$ are mutually exclusive and $P(S_i) = 1/2^9$ for all $i$. Then by (a), we have $P(S) = \sum_{i=1}^{10} P(S_i) = 10/2^9 = 5/256$.

2. Let $X$ and $Y$ be independent random variables from $N(0, 1)$. (a) Find the probability density function of $|X|$. (b) Find the probability density function of $|X/Y|$.

解答:

(a) Since $X \sim N(0, 1)$, we have that for any $x \geq 0$,

$$P(|X| \leq x) = P(-x \leq X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-t^2/2} \, dt = \frac{2}{\sqrt{2\pi}} \int_{0}^{x} e^{-t^2/2} \, dt.$$ 

Hence the probability density function of $|X|$ is $2 \exp\{-x^2/2\}/\sqrt{2\pi}$, where $x \geq 0$.

(b) For any $z \geq 0$, since $X$ and $Y$ are independent and $|Y|$ has probability density function $f_{|Y|}(y) = 2 \exp\{-y^2/2\}/\sqrt{2\pi}$, $y \geq 0$, we have

$$P(|X/Y| \leq z) = \int_{0}^{\infty} P(|X| \leq zy)f_{|Y|}(y) \, dy = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} P(|X| \leq zy)e^{-y^2/2} \, dy.$$ 

Since $|X|$ also has probability density function $f_{|X|}(x) = 2 \exp\{-x^2/2\}/\sqrt{2\pi}$, $x \geq 0$, we see that

$$\frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} P(|X| \leq zy)e^{-y^2/2} \, dy = \frac{2}{\pi} \int_{0}^{\infty} \int_{0}^{\infty} e^{-(x^2+y^2)/2} \, dx \, dy.$$ 

Then $|X/Y|$ has probability density function

$$f_{|X/Y|}(z) = \frac{2}{\pi} \int_{0}^{\infty} y e^{-(z^2y^2+y^2)/2} \, dy = \frac{-2}{\pi(z^2+1)} e^{-y^2(z^2+1)/2}|_{y=0}^{\infty} = \frac{2}{\pi(z^2+1)}.$$
where \( z \geq 0 \).

3. If \( X \) and \( Y \) are independent binomial random variables with identical parameters \( n \) and \( p \), find the conditional probability mass function of \( X \), given that \( X + Y = m \). State the name of the conditional distribution.

解答:

\[
P\{X = k|X + Y = m\} = \frac{P\{X = k, X + Y = m\}}{P\{X + Y = m\}}
= \frac{P\{X = k, Y = m - k\}}{P\{X + Y = m\}}
= \frac{\binom{n}{k} p^k (1 - p)^{n-k}\binom{m}{m-k} p^{m-k} (1 - p)^{n-m+k}}{\binom{2n}{m} p^m (1 - p)^{2n-m}}
= \frac{\binom{n}{k} \binom{n}{m-k}}{\binom{2n}{m}}
\]

\( X|X + Y = m \sim \) hypergeometric distribution\((2n, n, m)\).

4. Suppose that the random variable \( Y_1, Y_2, \ldots, Y_n \) satisfy

\[
Y_i = \beta x_i + \epsilon_i \quad i = 1, \ldots, n,
\]

where \( x_1, \ldots, x_n \) are fixed constants, and \( \epsilon_1, \epsilon_2, \ldots, \epsilon_n \) are iid \( N(0, \sigma^2) \), \( \sigma^2 \) unknown.

(a) Find a two-dimensional sufficient statistic for \((\beta, \sigma^2)\).

(b) Find the MLE of \( \beta \), and show that it is an unbiased estimator of \( \beta \).

(c) Find the distribution of the MLE of \( \beta \).

解答:

(a)

\[
L(\theta|y) = \prod_i \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{1}{2\sigma^2} (y_i - \beta x_i)^2 \right)
= (2\pi \sigma^2)^{-n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_i (y_i^2 - 2\beta x_i y_i + \beta^2 x_i^2) \right)
= (2\pi \sigma^2)^{-n/2} \exp \left( -\frac{\beta^2}{2\sigma^2} \sum_i x_i^2 \right) \exp \left( -\frac{1}{2\sigma^2} \sum_i y_i^2 + \frac{\beta}{\sigma^2} \sum_i x_i y_i \right).
\]

By Theorem 6.1.2, \((\sum_i Y_i^2, \sum_i x_i Y_i)\) is a sufficient statistic for \((\beta, \sigma^2)\).

(b)

\[
\log L(\beta, \sigma^2|y) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_i y_i^2 + \frac{\beta}{\sigma^2} \sum_i x_i y_i - \frac{\beta^2}{2\sigma^2} \sum_i x_i^2.
\]
For a fixed value of $\sigma^2$,

$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma^2} \sum_i x_i y_i - \frac{\beta}{\sigma^2} \sum_i x_i^2 = 0 \quad \Rightarrow \quad \hat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}.$$  

Also,

$$\frac{\partial^2 \log L}{\partial \beta^2} = \frac{1}{\sigma^2} \sum_i x_i^2 < 0,$$

so it is a maximum. Because $\hat{\beta}$ does not depend on $\sigma^2$, it is the MLE. And $\hat{\beta}$ is unbiased because

$$E[\hat{\beta}] = \frac{\sum_i x_i E[Y_i]}{\sum_i x_i^2} = \frac{\sum_i x_i \cdot \beta x_i}{\sum_i x_i^2} = \beta.$$  

(c) $\hat{\beta} = \sum_i a_i Y_i$, where $a_i = x_i / \sum_j x_j^2$ are constants. By Corollary 4.6.10, $\hat{\beta}$ is normally distributed with mean $\beta$, and

$$\text{Var}(\hat{\beta}) = \sum_i a_i^2 \text{Var}(Y_i) = \sum_i \left(\frac{x_i}{\sum_j x_j^2}\right)^2 \sigma^2 = \frac{\sum_i x_i^2}{(\sum_j x_j^2)^2} \sigma^2 = \frac{\sigma^2}{\sum_i x_i^2}.$$  

5. Consider a random sample $\{X_1, X_2, \ldots, X_n\}$ from a Pareto distribution. The CDF of a Pareto distribution is $F(x; \vartheta) = 1 - (1 + x)^{-\vartheta}, x > 0, \vartheta > 0$.

(a) Show that $-2 \sum_{i=1}^n \ln[1 - F(X_i; \vartheta)] \sim \chi^2(2n)$ has a chi-squared distribution with two degrees of freedom.

(b) Use the result of (a) to find a $100(1 - \alpha)$% confidence interval of the parameter $\vartheta$.

解答:

(a) Since $F(x; \vartheta)$ is the CDF of $X_i$, $1 - F(X_i; \vartheta) \sim U(0, 1)$, and consequently $-\ln(1 - F(X_i; \vartheta)) \sim \exp(1)$ and $-2 \sum_{i=1}^n \ln[1 - F(X_i; \vartheta)] \sim \chi^2(2n)$.

(b) $\left(\frac{x_i^2 / \alpha(2n)}{2 \sum_{i=1}^n \ln(1+x_i)}, \frac{x_i^2 / \alpha(2n)}{2 \sum_{i=1}^n \ln(1+x_i)}\right)$

6. Let $X_1, X_2, \ldots, X_n$ be independent r.v.'s with p.d.f. $f$ given by

$$f(x; \vartheta) = \frac{1}{\vartheta} e^{-x/\vartheta} I_{(0, \infty)}(x), \quad \vartheta \in \Omega = (0, \infty)$$

Derive the UMP (uniformly most powerful test) for testing the hypothesis $H_0 : \vartheta \geq \vartheta_0$ against the alternative $\vartheta < \vartheta_0$ at level of significance $\alpha$.

解答：Since the family of p.d.f. $\{f(x; \vartheta), \vartheta \in \Omega\}$ is a one-parameter exponential family, it has the MLR (monotone likelihood ratio) property in $V$, where $V(X_1, x_2, \ldots, x_n) = \sum_{j=1}^n x_j$. The UMP test rejects $H_0$ if $\sum_{i=1}^n x_j < C, C : P_{\vartheta_0}(\sum_{i=1}^n X_j < C) = \alpha$. Since $\sum_{i=1}^n X_j \sim \text{Gamma}(n, \vartheta), C$ is the $\alpha^{th}$ percentile of the Gamma$(n, \vartheta)$ distribution.