Exercise 4 Compulsory

Problem 1. Foundation of Statistics

(a) To prove linearity of the expected value of simple variables X and Y, one have to prove that:

$$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y].$$

The definition of the expected value of a simple random variable is:

$$E[X] = \sum_{i=1}^{N} x_i P(X = x_i)$$
$$= \sum_{i=1}^{N} X(\omega) P(\omega)$$

Here, $X(\omega) = x_i$, $\omega \in \Omega$ is an event.

(i) Proof using two Bernoulli variables X and Y

$$X(\omega) = \begin{cases} 0, & \omega \in A^c \\ 1, & \omega \in A \end{cases}$$
$$Y(\omega) = \begin{cases} 0, & \omega \in B^c \\ 1, & \omega \in B \end{cases}$$
$$(X+Y)(\omega) = \begin{cases} 0, & \omega \in (A^c \cap B^c) \\ 1, & \omega \in (A \cap B^c) \cup (A^c \cap B) \\ 2, & \omega \in (A \cap B) \end{cases}$$

$$E[X+Y] = 0 \cdot P(A^c \cap B^c) + 1 \cdot P(A \cap B^c) + 1 \cdot P(A^c \cap B) + 2 \cdot P(A \cap B)$$
$$= P(A \cap B^c) + P(A^c \cap B) + P(A \cap B) + P(A \cap B)$$
$$P(A \cap B^c) = P(A) - P(A \cap B)$$
$$P(A^c \cap B) = P(B) - P(A \cap B)$$
$$\Rightarrow E[X+Y] = P(A) + P(B)$$
$$= 0 \cdot P(A) + 1 \cdot P(A) + 0 \cdot P(B) + 1 \cdot P(B)$$
$$= E[X] + E[Y]$$

$$E[\alpha X] = \sum_{\omega} (\alpha X)(\omega) P(\omega)$$
$$= \sum_{i=1}^{n} \alpha x_i P(X = x_i)$$
$$= \alpha E[X]$$

(ii) A simple variable and a Bernoulli variable

(iii) Two simple random variables

$$X = \sum_{\omega} X(\omega) \mathbb{1}(\omega)$$
$$= \sum_{i=1}^{n} x_i \mathbb{1}(X = x_i)$$
$$Y = \sum_{\omega} Y(\omega) \mathbb{1}(\omega)$$
$$= \sum_{j=1}^{m} y_j \mathbb{1}(Y = y_i)$$
$$X + Y = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j) \mathbb{1}(X = x_i \cap Y = y_j)$$
$$E[X+Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j) P(X = x_i \cap Y = y_j)$$
$$= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i P(X = x_i \cap Y = y_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} y_j P(X = x_i \cap Y = y_j)$$
$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} P(X = x_i \cap Y = y_j) + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} P(X = x_i \cap Y = y_j)$$
$$\sum_{j=1}^{m} P(X = x_i \cap Y = y_j) = P(X = x_i)$$
$$\sum_{i=1}^{n} P(X = x_i \cap Y = y_j) = P(Y = y_j)$$
$$\Rightarrow E[X+Y] = E[X] + E[Y]$$
$$E[\alpha X] = \sum_{i=1}^{n} \alpha x_i P(X = x_i)$$
$$= \alpha \sum_{i=1}^{n} x_i P(X = x_i)$$
$$= \alpha E[X]$$

(b) Proof of $E[\varphi(W)] = E_W(\phi)$:

We have $E[\varphi(W)] = \sum_{\omega} \varphi(W(\omega)) P(\varphi(W) = \varphi(W(\omega)))$

$$= \sum_{w} \varphi(w) P(\varphi = \varphi(w)), \quad w = W(\omega)$$
$$= E_W[\varphi]$$

Alternative proof of linearity:

$$\phi(x,y) = \alpha x + \beta y$$
$$E[\phi(X,Y)] = E[\alpha X + \beta Y]$$
$$= \sum_{\omega \in \Omega} (\alpha X + \beta Y)(\omega) P(\alpha X + \beta Y = (\alpha X + \beta Y)(\omega))$$
$$= \sum_{\omega \in \Omega} \alpha X(\omega) P(\omega) + \sum_{\omega \in \Omega} \beta Y(\omega) P(\omega)$$
$$= \alpha E[X] + \beta E[Y]$$

(c) Since a statistic is a function of the data, the level set of the data is included in the level set of the statistic

The likelihood statistic is the minimal sufficient statistic.

Since the likelihood statistic is a sufficient statistic, the inference will be the same about θ, if $T(x) = T(y)$, whether $X = x$ or $Y = y$ is used

Problem 2. Statistics and Darts

(a) We have $\Omega_D = \{d \mid d = (x,y),\ x,y \in \mathbb{R} \cdot \text{meter}\}$

Let B denote the whole board

Let B_s denote the s-point ring

$$A_0 = \{d \mid d \notin B\}$$
$$A_s = \{d \mid d \in B_s\}, \quad s = 1, \dots, 9$$
$$A_{20} = \{d \mid d \in B \setminus \{B_s\}\}, \quad s = 1, \dots, 9$$

(b) We have 11 different $(D \in A_s)$, and either $(D \in A_s)$ is true or it is false.

Therefore we have 2^{11} different events.

Kolmogorov axioms say for a family $\mathcal{E}_D$ of events

(i) $\emptyset \in \mathcal{E}_D$

(ii) $A \in \mathcal{E}_D \Rightarrow A^c \in \mathcal{E}_D$

(iii) $A_0, A_1, \dots, A_9, A_{20} \Rightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{E}_D$

Obviously, the empty set is in $\mathcal{E}_D$ because if we don't throw the dart, nothing happens.

Obviously, the complement for an event A is also in $\mathcal{E}_D$, because if we get A_0 then A_0^c is just any other outcome.

Also, the union of the sets is in $\mathcal{E}_D$, because all subsets of Ω_D are in $\mathcal{E}_D$

The definition of a sigma algebra is just the collection of subsets of a sample space

Since $\mathcal{E}_D$ is defined using subsets of Ω_D, Ω_D is a sample space.

Also, $P(\Omega_D) = 1$, $P(A_s) \geq 0$, $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Since our events are the events of hitting A_s, and not every point, every subset of Ω_D are not events.

(c) $\Omega_\theta = \mathbb{R}^n$ can be used as a model space, because we have 11 parameters.

The $R(\theta)$ must be such that $\sum_{i=1}^{n} p_i = 1$

(d) The score S is a statistic because it is given by the data, i.e. the points.

Since S is a simple function, $\mu = E[S] = \sum_{i=1}^{n} s_i P(S = s_i)$, i.e. it is a value calculated from the probability and the scores obtainable.

The law of large numbers says the average approaches μ after many trials, i.e. $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{s_i}{n} = \mu$

This tells us that $\lim_{n\to\infty} \sum_{i=1}^{n} \frac{s_i}{n} = \sum_{i=1}^{n} s_i P(S = s_i)$

Which says that $P(S = s_i) = \frac{1}{n}$

(e) The data space Ω_D then becomes $\Omega_D = \mathbb{R}^n$

The experiment is the result of a single dart throw.

$\mathcal{E}_D$ then has 2^{11n} members.

Now $\{d\}$ becomes an event because it is in the data space and event space.

(f) m_s defines a random variable M_s because it is a function $M_s: \Omega \to \mathbb{R}$,

and a random vector $M = (M_1, \dots, M_{20})$ because $M: \Omega \to \mathbb{R}^{11}$.

M belongs to the multinomial distribution, and is therefore also in the exponential family of distributions.

The complete statistic in the exponential family is $T(X) = \left(\sum_{i=1}^{n} t_1(X_i), \dots, \sum_{i=1}^{n} t_s(X_i)\right)$.

In our case, $M = \left(\sum_{i=1}^{n} m_1(d_i), \dots, \sum_{i=1}^{n} m_{20}(d_i)\right)$, so it is complete and minimal sufficient statistic.

(g)
$$L(p) = n! \prod_s \frac{p_s^{m_s}}{m_s!}$$
$$\ell(p) = \log(n!) + \log\left(\prod_s \frac{p_s^{m_s}}{m_s!}\right)$$
$$= \log(n!) + \sum_s m_s \log(p_s) - \sum_s \log(m_s!)$$

Lagrange multiplier λ

$$\mathcal{L}(p,\lambda) = \ell(p) + \lambda\Big(1 - \underbrace{\sum_s p_s}_{=1}\Big)$$
$$\frac{\partial}{\partial p_s} \mathcal{L}(p,\lambda) = \frac{m_s}{p_s} - \lambda$$
$$= 0$$

$$\frac{m_s}{p_s} = \lambda$$
$$\lambda p_s = m_s$$
$$\lambda \sum_s p_s = \sum_s m_s = n$$
$$\Rightarrow \lambda = n$$
$$\Rightarrow \hat{p}_s = \frac{m_s}{n} \quad \forall s$$
$$\Rightarrow \hat{\theta} = \left(\frac{m_0}{n}, \frac{m_1}{n}, \dots, \frac{m_9}{n}, \frac{m_{20}}{n}\right)$$
$$E[\hat{p}_s] = E\left[\frac{m_s}{n}\right]$$
$$= \frac{1}{n} E[m_s]$$
$$= \frac{1}{n} \cdot n p_s$$
$$= p_s \text{ unbiased}$$
$$\hat{\mu} = E[S]$$
$$= \sum_{i=1}^{11} s_i P(S = s_i)$$
$$= \sum_{i=1}^{11} s_i \hat{p}_i$$
$$E[\hat{\mu}] = \sum_{i=1}^{11} s_i E[\hat{p}_i]$$
$$= \sum_{i=1}^{11} s_i p_i$$
$$= \mu \text{ unbiased}$$

(h) The data space now becomes $\Omega_D \subset \mathbb{R}^{2 \times n}$

Sigma algebra $\mathcal{E}_D$ becomes $\mathcal{B}(\mathbb{R}^{2 \times n})$

$$L(x,y \mid \mu,\sigma) = \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{1}{2}\frac{x_i^2}{\sigma^2}} \cdot e^{-\frac{1}{2}\frac{y_i^2}{\sigma^2}}$$
$$= \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{x_i^2 + y_i^2}{2\sigma^2}}$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^n \cdot e^{-\sum_{i=1}^{n} \frac{x_i^2 + y_i^2}{2\sigma^2}}$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^n \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2 + y_i^2)}$$
$$\ell(x,y \mid \mu,\sigma) = n\log\left(\frac{1}{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2 + y_i^2)$$
$$= -n\log(2\pi) - n\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i^2 + y_i^2)$$
$$\frac{\partial \ell}{\partial \sigma} = -\frac{2n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i^2 + y_i^2) = 0$$
$$\Rightarrow -2n\sigma^2 + \sum_{i=1}^{n}(x_i^2 + y_i^2) = 0$$
$$\Rightarrow \sigma^2 = \frac{1}{2n}\sum_{i=1}^{n}(x_i^2 + y_i^2)$$
$$\Rightarrow \hat{\sigma} = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}(x_i^2 + y_i^2)}$$

Since $\binom{x}{y} \sim N\left(\binom{0}{0}m, \sigma I\right)$

$$f(x,y \mid \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$
$$= \frac{1}{2\pi} e^{-\ln(\sigma^2)} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}}$$
$$= \frac{1}{2\pi} e^{-\ln(\sigma^2) - \frac{x^2}{2\sigma^2} - \frac{y^2}{2\sigma^2}}$$
$$f(x,y \mid \hat{\sigma}) = \frac{1}{2\pi} e^{-\ln\left(\frac{1}{2n}\sum_{i=1}^{n}(x_i^2+y_i^2)\right) - \frac{nx^2}{\sum_{i=1}^{n}(x_i^2+y_i^2)} - \frac{ny^2}{\sum_{i=1}^{n}(x_i^2+y_i^2)}}$$
$$=$$

(i)

$$\mu^* = E[\hat{\mu}(D) \mid \hat{\sigma}(D) = \hat{\sigma}]$$
$$E[\mu^*] = E[E[\hat{\mu}(D) \mid \hat{\sigma}(D) = \hat{\sigma}]]$$
$$= E[\hat{\mu}(D)], \quad \hat{\mu} \text{ unbiased}$$
$$= \mu \text{ unbiased}$$

From Rao-Blackwell we have

$$Var[T^*] \leq Var[T], \quad T^* = E[T \mid S]$$

where S is complete and sufficient.

So $Var[\mu^*] \leq Var[\hat{\mu}]$ since $\hat{\sigma}$ is complete minimal sufficient statistic.

(j)
$$\hat{\sigma}(gd) = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}\left((gx_i)^2 + (gy_i)^2\right)}$$
$$= \sqrt{\frac{g^2}{2n}\sum_{i=1}^{n}(x_i^2 + y_i^2)}$$
$$= |g|\sqrt{\frac{1}{2n}\sum_{i=1}^{n}(x_i^2 + y_i^2)}$$
$$= g\hat{\sigma}(d), \quad g \text{ is positive}$$

So $\hat{\sigma}$ is equivariant.

(k)
$$\hat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n}(x_i^2 + y_i^2)$$
$$x_i^2 \sim N(0,\sigma^2)$$
$$y_i^2 \sim N(0,\sigma^2)$$
$$\Rightarrow \sum_{i=1}^{n}(x_i^2 + y_i^2) \sim \chi_{2n}^2$$
$$cX \sim \Gamma(k = v/2, \theta = 2c)$$
$$X \sim \chi_v^2$$
$$\Rightarrow \hat{\sigma}^2 \sim \Gamma\left(k = n, \theta = \frac{1}{n}\right)$$
$$\Gamma(\alpha = n, \beta = n)$$
$$\Rightarrow |\hat{\sigma}| \sim \chi_n \quad \text{Chi-distribution)}$$

We know that when $|X| \sim \chi_k$ with k degrees of freedom, then

$$X \sim N(0,1)$$

So

$$\hat{\mu} = \bar{X} \pm z_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} \text{ C.I.}$$

and

$$\hat{\sigma} =$$

Let $a = \frac{(x,y)}{(\bar{x}+\bar{y})}$

$$\Rightarrow a = \left(\frac{x}{\bar{x}+\bar{y}}, \frac{y}{\bar{x}+\bar{y}}\right)$$
$$x \sim N(0,\sigma^2) \Rightarrow \frac{x}{\bar{x}+\bar{y}} \sim N\left(0, \frac{\sigma^2}{(\bar{x}+\bar{y})^2}\right)$$
$$y \sim N(0,\sigma^2) \Rightarrow \frac{y}{\bar{x}+\bar{y}} \sim N\left(0, \frac{\sigma^2}{(\bar{x}+\bar{y})^2}\right)$$

Basu's theorem:

Since a is ancillary, $\hat{\sigma}$ is independent of a.