① The definition: $EX = \sum_i^n x_i P(A_i)$ where $A_i = (X = x_i)$. This gives more generally (since P is finitely additive)

$$E\Big(\sum_i x_i A_i\Big) = \sum_i^n x_i P(A_i) \qquad (1)$$

as long as $A_i \cap A_j = \emptyset$ $i \neq j$ and $\Omega = A_1 \cup \dots \cup A_n$.

a) (i) $E(\alpha X + \beta Y) = \alpha P(XY^c) + (\alpha+\beta) P(XY)$
$+ \beta P(X^c Y) + 0 \cdot P(X^c Y^c)$

Ω

4 sets $A_1, \dots, A_n$

$\alpha EX = \alpha \cdot [P(XY) + P(XY^c)]$

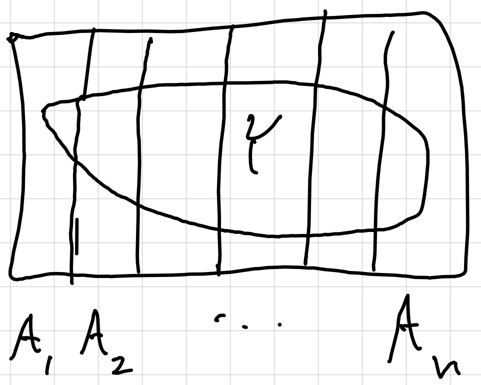
$\beta EY = \beta\, [P(YX) + P(YX^c)]$

$\Rightarrow E(\alpha X + \beta Y) = \alpha EX + \beta EY$

ii) $E(\alpha X + \beta Y) = \sum_i (\alpha x_i P(A_i Y^c) + (\alpha x_i + \beta) P(A_i Y))$

$\alpha EX = \sum_i \alpha x_i P(A_i)$, $\beta EY = \beta P(Y)$, $A_i = A_i Y^c + A_i Y$

ii) $Y = \sum_i (YA_i + YA_i^c)$ and insertion gives $E(\alpha X + \beta Y) = \alpha EX + \beta EY$



$2n$ pieces.

iii) Let X take n values and Y take m values, so Ω is divided in $n \times m$ pieces (some may be $= \emptyset$). The value of $\alpha X + \beta Y$ is $\alpha x_i + \beta y_j$ on $A_{ij} = (X = x_i) \cap (Y = y_j)$.

The value of αX is αx_i on A_{ij}. Value of βY is βy_i on A_{ij}

$$E(\alpha X + \beta Y) = \sum_{i,j} (\alpha x_i + \beta y_i) P(A_{ij})$$

$$= \alpha \sum_{i,j} x_i P(A_{ij}) + \beta \sum_{i,j} y_j P(A_{ij})$$

$$= \alpha EX + \beta EY$$

① b) $E\varphi(W) = \sum_i \varphi_i \cdot P(W = w_i)$

with $\varphi_i = \varphi(w_i)$.

$E_W \varphi = \sum_i \varphi_i P_W(\{w_i\})$

so $E\varphi(W) = E_W \varphi$.

Note: Eq. 1 is used in both cases: Some of the φ_i values may be equal. In the first case it is used on Ω. In the second case it is used on Ω_W.

1a&b short $ET = \sum_t t f(t) \overset{\nabla}{=} \sum_x t(x) f(x)$

since $f(t) = \sum_{t(x)=t} f(x)$ from additivity of P.

$E(\alpha X + \beta Y) = \sum_{x,y} (\alpha x + \beta y) f(x,y) = \alpha EX + \beta EY$.

$E\varphi(W) = E_W \varphi$ since $P_{\varphi(W)}(A) = (P_W)_\varphi(A)$.

(1) b) Let $w = (x, y)$, so

$$E(\alpha X + \beta Y) = E\varphi(W)$$

$$= E_W \varphi = \sum_w \varphi(w) P_W\{w\}$$

$$= \alpha \sum_w x P_W\{w\} + \beta \sum_w y P_W\{w\}$$

$$= \alpha EX + \beta EY$$

using the change-of-variables thm again on $(x, y) \mapsto x$ and $(x, y) \mapsto y$. Easier proof by first proving

$$E\varphi(W) = \sum_w f(w)\varphi(w)$$

directly:

$$E\varphi(W) = \sum_y y P(\varphi(W) = y) = \text{etc.}$$

a) e I) with induction

①

$C_n : E\left(\sum_{i=1}^{n} \alpha_i A_i\right) = \sum_{i=1}^{n} \alpha_i EA_i$

$n=1 : E\alpha A = \sum_i \alpha a_i EA_i = \alpha EA$

$n=k+1 : E\left(\sum_{i=1}^{k} \alpha_i A_i + \alpha_{k+1} A_{k+1}\right)$

$\overset{(ii)}{=} E\left(\sum^{k} \alpha_i A_i\right) + \alpha_{k+1} EA_{k+1}$

$\overset{C_k}{=} \sum_i^{k+1} \alpha_i EA_i$

so C_n is true for $n = 1, 2, 3, \cdots$.

iii) : $E(\alpha X + \beta Y)$

$= E\left(\sum \alpha x_i X_i + \sum \beta y_j Y_j\right)$

$= \sum \alpha x_i EX_i + \sum \beta y_j Y_j$

$= \alpha EX + \beta EY$

1 c)

$$\begin{tikzcd}
\Omega \arrow[r, "T"] \arrow[d, "X"'] & \Omega_T \\
\Omega_X \arrow[ur, "t(\cdot)"'] &
\end{tikzcd}$$

$$(X = x) \subset (T = t)$$

when $t(x) = t$. It is assumed that $D(X) = D(T)$.

The inclusion of level sets is equivalent with $T = t(X)$, so equality of level sets means that $t(\cdot)$ is a one-one transformation between $R(X)$ and $R(T)$.

The factorization theorem proves that L is minimal, and hence its level sets equals the level sets of any minimal suff. stat.

(1) c) Suff. principle: If $t(x)=t(\tilde{x})$ then same inference for x and $\tilde{x}$. (This must hold for any sufficient statistic $T=t(X)$. It is hence ensured if it holds for a single minimal sufficient statistic.)

Suff. principle: Same inference if $x \sim \tilde{x}$

The level sets of $t(\cdot)$ defines a partition $\{A_t\}$ $t \in R(T)$ by
$$A_t = \{x \mid t(x) = t\}.$$
of $R(X)$. This defines an equivalence relation (any partition does!)

$$x \sim \tilde{x} \iff t(x) = t(\tilde{x})$$

Reflexive. i) $x \sim x$ Symmetric ii) $x \sim y \iff y \sim x$ Transitive iii) $x \sim y, y \sim z \Rightarrow x \sim z$

$$A_{20} = \{ d \mid |d| < R_{20} \}$$

2

a) $$A_s = \{ d \mid R_{s+1} \leq |d| < R_s \} \quad 0 \leq s \leq 9$$

$$0 < R_{20} < R_9 < \cdots < R_1 < R_0 = \infty$$

$R_{10} = R_{20}$. $\quad s \in R(S) = \{0, 1, \cdots, 9, 20\}$

b) Let $I \subset R(S)$ and define $A_I = \bigcup_{s \in I} A_s$. The family

$$\mathcal{E}_D = \{ A_I \mid I \subset R(S) \}$$

has 2^{11} elements since $2^{11} = |2^{R(S)}|$ = number of subsets of $R(S)$ = number of binary sequences of length 11.

$$P_D(A_I) \overset{A!}{=} \sum_{s \in I} P_D(A_s)$$

(2) b) Check of axioms for $\mathcal{E}_D$:

i) $\emptyset \in \mathcal{E}_D$ since $A_\emptyset = \emptyset$ //

ii) $(A_{\underline{I}})^c = A_{\underline{I}^c}$ //

iii) $A_{\underline{I}_1} \cup A_{\underline{I}_2} \cup \cdots = A_{\underline{I}_1 \cup \cdots}$ //

Check of axioms for P_D:

(i) $P_D(A_{\underline{I}}) = \sum_{s \in \underline{I}} P_D(A_s) \geq 0$ //

(ii) $P_D(\Omega_D) = P_D(A_{R(S)})$

$= \sum_{s \in R(S)} P_D(A_s) = 1$ //

(iii) $A_{\underline{I}_1}, A_{\underline{I}_2}, \ldots$ disjoint

$P_D(\cup A_{\underline{I}_i}) = \sum_i P(A_{\underline{I}_i})$ //

b) Alternative and better:

(2)

$$(D \in A_s) = \{\omega \mid D(\omega) \in A_s\}$$

is by assumption an event for all s. This gives that

$$(D \in A_I) = \bigcup_{s \in I} (D \in A_s) \text{ is}$$

always an event. It follows then that P_D defined by $P_D(A) = P(D \in A)$ is a probability since P is.

Ω_D equipped with $\mathcal{E}_D$ is a sample space.

Ω_D with P_D is a probability space.

$\{(0,0)\} \subset \Omega_D$, but is not an event

b) since $\{(0,0)\} \notin \mathcal{E}_D$.

(2) More generally: Any set $A \subset \Omega_D$ with $A \notin \mathcal{E}_D$ is not an event. In particular any proper non-empty of any A_s.

Many subsets of Ω_D are not events.

Note: This is the most common situation. An event is a set, but a set need not be an event.

c) $\Theta = (p_0, \dots, p_9, p_{10})$

$\in \Omega_\Theta = \mathbb{R}^{11}$.

$$R(\Theta) = \{\theta \mid 0 \leq \theta_s \leq 1, \sum_s \theta_s = 1\}$$

② d) $s(d) = \sum_{s \in R(S)} s \cdot A_s(d)$ so

s is a function of the data $\equiv$ A statistic

Level set

$$A_s = \{d \mid \varphi(d) = s\}$$

from $\varphi(d) = \sum_{s \in R(S)} s A_s(d)$

$$S = \varphi(D) \;\overset{?}{=}\; s(D)$$

$\theta_s = p_s$

$\mu = ES = \sum s f(s) = \sum s \theta_s$ is a function of θ = A parameter

$= E_D \varphi$ by Change-of. Var.

2

d) Law of large numbers

$$E(S) = \lim_{n\to\infty} \frac{S_1 + \cdots + S_n}{n}$$

if $S_i \sim S$ and independent.

Interpretation: If the dart exp. is repeated many times, then the empirical average $\bar{S}$ tends to $E(S)$.

It must be assumed that $E|S| < \infty$, but this is fine here since $|S| \leq 20$.

$$P(S \in A) = \lim_{n\to\infty} \frac{(S_1 \in A) + \cdots + (S_n \in A)}{n}$$

so relative frequency $\to$ probability.

(2) e) $\Omega_\Theta = \mathbb{R}^{11}$ still.

$$\Omega_D = \{d \mid d_i \in \mathbb{R}^2 \cdot \mathrm{lm},\ i=1,\dots,n\}$$
$$= \mathbb{R}^{2n} \cdot \mathrm{lm}$$

$\{d\} \notin \mathcal{E}_D$ so $\{d\}$ is not an event.

The previous partition in 11 events gives now a partition into 11^n events. The number of events in $\mathcal{E}_D$ is hence

$$|\mathcal{E}_D| = 2^{(11^n)}$$

Experiment = n throws with a dart.

(2) f) $M_s = \sum_{i=1}^{n} A_s(D_i)$ is a random variable since it is a sum of random variables = # darts hitting A_s

$\sim B(n, p_s)$

$M \sim \text{Multinomial}(n, \theta)$

it is minimal and complete, and the distribution is from an exponential family.

Argument: Consider $n=1$ first. The distribution of D_1 is in one-one correspondence with the distribution of $[A_0(D_1), \dots, A_{20}(D_1)]$. Similarly for D and $[A_0(D), \dots, A_{20}(D)]$

② f) The likelihood for the statistic is

$$L = \prod_{i=1}^{n} p_0^{x_0(i)} \cdots p_{20}^{x_{20}(i)}$$

where $x_s(i) = A_s(d_i)$

= indicator for dart i hitting A_s.

$$L = e^{m_0 \ln p_0 + \cdots + m_{20} \ln p_{20}}$$

is an exp. family. Note restriction $m_0 + \cdots + m_{20} = n$ and $p_0 + \cdots + p_{20} = 1$.

This proves minimal completeness (Can change coordinates to 10 dimensional to get stand. form.)

(2) g) $\hat{\theta} = \frac{m}{n}$ (Elimination of p_{20} or Lagrange multiplier for $\hat{\theta} = \arg\min L$ with $\sum_s \theta_s = 1$.)

$\mu = \sum_s s p_s$, $\hat{\mu} = \sum_s s \hat{\theta}_s$

Yes, unbiased since

$E A_s(D_i) = p_s$

and $E \hat{\mu}(D) = \sum_s s E \hat{\theta}_s(D) = \sum_s s \theta_s = \mu$

h) The set Ω_D is unchanged, but the sample space Ω_D is different $= \mathbb{R}^{2n} \cdot \mathrm{Im}$ with $\mathcal{E}_D = $ Borel

(2) h) σ-algebra = smallest family of events containing all $(-\infty, a]$, $a \in \mathbb{R}^{2n}$. Im.

$$L = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\,\sigma}\right]^2 e^{-\frac{x_i^2+y_i^2}{2\sigma^2}}$$

$$= \left[\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma}\right]^{2n} \cdot e^{-\frac{\sum^n r_i^2}{2\sigma^2}}$$

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(x_i^2+y_i^2)}{2n}}$$

Complete minimal sufficient since exp. family form

i) $\hat{\mu}(d)$ cannot be computed from the (minimal) suff. stat. $\hat{\sigma}$, so it conflicts with the suff. principle

$\mu(\hat{\sigma})$ is in harmony with

② i) Note: $\mu = \mu(\alpha)$ since each $p_s = p_s(\alpha)$ by integration over A_s.

μ^* is unbiased since $\hat{\mu}(\alpha)$ is unbiased

$$\left(E(E(X \mid T)) = E(X) \right)$$

It is an estimator since $\hat{\alpha}$ is suff., and optimal unique from completeness of $\hat{\alpha}$

$\Rightarrow$ μ^* is the unique UMVU.

(Rao-Blackwellization, Leh.-Scheffe)

2) i) Std $\mu^*(D)$ depends on α.

u can be calculated by simulation using $\hat{\alpha}$ as α by giving $\mu^*_1, \ldots, \mu^*_N$ ($N = 10^6$) giving S_N as u. This requires calculation of $\mu^*(\alpha)$ which also can be done by simulation (but actually also analytically)

Calculation of both μ^* and its standard uncertainty is challenging.

(2) i) $g(D)$ bivariate normal sample with standard dev. $g\sigma$): Group model

Family of distributions for data D is invariant when G acts on the data.

(*) $x_i = \sigma u_i$, $y_i = \sigma v_i$
i.i.d. $u_i, v_i \sim N(0,1)$.

$\hat{\sigma} = \hat{\sigma}(d)$ and

$\hat{\sigma}(gd) = g\hat{\sigma}(d)$ from (*) and formula for $\hat{\sigma}$.

Is equivariant

(2) b) $2n\hat{\sigma}^2/\sigma^2 = \chi^2_{2n}$

$\Rightarrow \hat{\sigma}^2 \sim \frac{\sigma^2}{2n}\chi^2_{2n}$

(equal Gamma also)

CI $1-\alpha = P(a \leq \chi^2_{2n} \leq b)$

$\frac{2n\hat{\sigma}^2}{\sigma^2}$

$\Rightarrow \frac{2n\hat{\sigma}^2}{b} \leq \sigma^2 \leq \frac{2n\hat{\sigma}^2}{a}$

$\left[\frac{1}{\sqrt{b}}, \frac{1}{\sqrt{a}}\right]\cdot\sqrt{2n}\,\hat{\sigma} = [\hat{\sigma}_1, \hat{\sigma}_2]$

② b) $\mu = \mu(\alpha)$ is decreasing in α (α small gives more high scores) so

$[\mu(\hat{\alpha}_2), \mu(\hat{\alpha}_1)]$

$\hat{\alpha}$ is complete suff. and

a is ancillary

$\left(a = \frac{(u, v)}{\bar{u} + \bar{v}} \right)$:-)

$\Rightarrow$ $\hat{\alpha}$ and a are independent

Orbits are rays, and each ray is uniquely given by a maximal invariant.

$(\hat{\alpha}, a) \underset{1-1}{\Leftrightarrow} d$ fails $\Rightarrow$ invariant. Not maximal invariant