Question (1973 STEP III Q10)

(i) The real numbers $a_1, \dots, a_n$ satisfy the constraint

$$\sum_{i=1}^{n} a_i = C,$$

where C is a given constant. Show that $\sum_{i=1}^{n} a_i^2$ is minimised subject to (*) by $a_i = C/n$ for $i = 1, \dots, n$. (ii) In an experiment to determine the mean body-weight μ of a species of moth, n moths of this species are weighed, and their weights $x_1, \dots, x_n$ recorded. It may be assumed that $x_1, \dots, x_n$ are uncorrelated and have common mean μ and common variance σ^2, where σ^2 is known. We wish to find the best linear unbiased estimator of μ, that is the function $\sum_{i=1}^{n} a_i x_i$ which has expectation μ and smallest variance. Assuming (i), find the appropriate values of the set $\{a_i\}$, and find the variance of the best linear unbiased estimator.

Question (1981 STEP III Q9)

Independent random variables $X_1, \dots, X_n$ have a uniform distribution on the interval $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$, where θ is unknown. Let

$$V = \max\{X_1, \dots, X_n\}, \quad U = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Show that if $\theta = \theta_0$,

$$P(V \leq x) = \left(x - \theta_0 + \frac{1}{2}\right)^n \quad \text{for } \theta_0 - \frac{1}{2} \leq x \leq \theta_0 + \frac{1}{2}.$$

Calculate the mean and variance of V and U. If $W = V - c$, where c is chosen so that $EW = \theta_0$, compare var U and var W. Suppose that n is large. Which of U, W would you use if you wished to estimate θ, and why?

$$\begin{aligned} P(V \leq x) &= P(\max(X_1, \dots, X_n) \leq x) \\ &= P(X_1 \leq x) \cdots P(X_n \leq x) \\ &= \frac{(x - (\theta_0 - \frac{1}{2}))}{1} \cdots \frac{(x - (\theta_0 - \frac{1}{2}))}{1} \\ &= \left(x - \left(\theta_0 - \frac{1}{2}\right)\right)^n \end{aligned}$$

$$\begin{aligned} \mathbb{E}(U) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \theta_0 \\ &= \theta_0 \end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(U) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) \\
&= \frac{1}{n^2}\frac{n}{12} \\
&= \frac{1}{12n}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(V) &= \int_{\theta_0-\frac{1}{2}}^{\theta_0+\frac{1}{2}} xn\left(x-\left(\theta_0-\frac{1}{2}\right)\right)^{n-1}\mathrm{d}x \\
&= n\int_0^1 \left(u+\left(\theta_0-\frac{1}{2}\right)\right)u^{n-1}\mathrm{d}u \\
&= n\left[\frac{u^{n+1}}{n+1}+\left(\theta_0-\frac{1}{2}\right)\frac{u^n}{n}\right]_0^1 \\
&= \frac{n}{n+1}+\theta_0-\frac{1}{2} \\
\mathrm{Var}(V) &= \mathbb{E}(V^2)-[\mathbb{E}(V)]^2 \\
&= \int_{\theta_0-\frac{1}{2}}^{\theta_0+\frac{1}{2}} x^2n\left(x-\left(\theta_0-\frac{1}{2}\right)\right)^{n-1}\mathrm{d}x-\left(\frac{n}{n+1}+\theta_0-\frac{1}{2}\right)^2 \\
&= n\int_0^1 \left(u+\theta_0-\frac{1}{2}\right)^2 u^{n-1}\mathrm{d}u-\left(\frac{n}{n+1}+\theta_0-\frac{1}{2}\right)^2 \\
&= n\left[\frac{u^{n+2}}{n+2}+(2\theta_0-1)\frac{u^{n+1}}{n+1}+\left(\theta_0-\frac{1}{2}\right)^2\frac{u^n}{n}\right]_0^1-\left(\frac{n}{n+1}+\theta_0-\frac{1}{2}\right)^2 \\
&= \frac{n}{n+2}+\frac{n}{n+1}(2\theta_0-1)+\left(\theta_0-\frac{1}{2}\right)^2-\left(\frac{n}{n+1}+\theta_0-\frac{1}{2}\right)^2 \\
&= \frac{n}{n+2}-\frac{n^2}{(n+1)^2} \\
&= \frac{n(n+1)^2-n^2(n+2)}{(n+2)(n+1)^2} \\
&= \frac{n^3+2n^2+n-n^3-2n^2}{(n+2)(n+1)^2} \\
&= \frac{n}{(n+2)(n+1)^2}
\end{aligned}$$

Therefore the variance of V is much smaller than U for large n, so we should choose that.

Question (1970 STEP IIII Q9)

If $x_1, x_2, \dots, x_n$ is a random sample from the uniform distribution with density function $f(x) = 1/\theta$, $0 < x < \theta$, where θ is an unknown parameter:

(i) find the maximum likelihood estimate $\hat{\theta}$ of θ,

(ii) find the density function of $\hat{\theta}$ and hence find its mean.

[The maximum likelihood estimate of a parameter θ based on a random sample $x_1, x_2, \dots, x_n$ from a distribution with density function $f(x, \theta)$ is that value of θ which maximizes the likelihood function $f(x_1, \theta)f(x_2, \theta)\dots f(x_n, \theta)$.]

1.

$$\begin{aligned} L(\theta) &= f(x_1, \theta)f(x_2, \theta)\dots f(x_n, \theta) \\ &= \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \geq \max(x_1, x_2, \dots) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

which is clearly maximised if $\theta = \max(x_1, x_2, \dots)$

2.

$$\begin{aligned} F_{\hat{\theta}}(t) &= P(\hat{\theta} < t) \\ &= P(\max(x_1, \dots) < t) \\ &= P(x_1 < t)P(x_2 < t)\cdots P(x_n < t) \\ &= \left(\frac{t}{\theta}\right)^n \\ &= \frac{t^n}{\theta^n} \end{aligned}$$

Therefore $f_{\hat{\theta}}(t) = n\theta^{-n}t^{n-1}$.

$$\begin{aligned} \mathbb{E}(\hat{\theta}) &= \int_0^\theta t f_{\hat{\theta}}(t) \mathrm{d}t \\ &= \int_0^\theta n\theta^{-n}t^n \mathrm{d}t \\ &= \frac{n}{n+1}\theta \end{aligned}$$

Question (1973 STEP IIII Q10)

X and Y are discrete valued random variables, and

$$\Pr(X = x, Y = y) = p(x, y), \quad \text{say.}$$

The expectation of X conditional on the value of Y being y is defined as $\mu(y)$, where

$$\mu(y) = E(X|Y = y) = \sum_x x \frac{p(x, y)}{b(y)},$$

and

$$b(y) = \Pr(Y = y),$$

so that

$$b(y) = \sum_x p(x, y).$$

Show that $E(X) = \sum \mu(y)b(y)$. By taking $Z = X^2$, find an expression for the variance of X in terms of $E(X|Y = y)$ and $E(X^2|Y = y)$. An ornithologist observes that the number of eggs laid by a sparrow in a nest is distributed approximately as a Poisson random variable with mean λ. He suspects that any egg has the same probability p of hatching, and that they are independent with respect to hatching. Denote by X the number of fledgelings from a nest, and denote by Y the number of eggs laid in that nest. Find expressions for

$$E(X|Y = y) \quad \text{and} \quad E(X^2|Y = y)$$

and hence find the (unconditional) mean and variance of X. A second ornithologist contests that the eggs in a nest are not independent with respect to hatching. He suspects that either, with probability π, the whole clutch of eggs hatches, or, with the probability $1 - \pi$, none of the clutch hatches. What are the mean and variance of X with this model? If you looked at a large sample of sparrows' nests, and found that the mean number of fledgelings per nest was 4, and the sample variance was 12, which ornithologist would you take to be more expert?

Question (1975 STEP IIII Q11)

The random variables $X_1, X_2, \ldots, X_n$ are independent and have identical probability distributions. The function ϕ of n arguments is such that $\phi(X_1, X_2, \ldots, X_n)$ has expectation μ and variance σ^2. Furthermore, ϕ is not symmetric, so that there is at least one pair of suffixes (i, j) such that with positive probability

$$\phi(X_1, \ldots, X_i, \ldots, X_j, \ldots, X_n) \neq \phi(X_1, \ldots, X_j, \ldots, X_i, \ldots, X_n).$$

The symmetrisation ψ of ϕ is defined by

$$\psi(X_1, \ldots, X_n) = \frac{1}{n!} \sum \phi(X_{i_1}, \ldots, X_{i_n})$$

where the summation is over all $n!$ permutations $(i_1, i_2, \ldots, i_n)$ of $(1, 2, \ldots, n)$. Prove that $\psi(X_1, X_2, \ldots, X_n)$ has expectation μ but variance less than σ^2. [A simpler version, using exactly the same strategy of proof, has $n = 2$.]

$$\begin{aligned}\mathbb{E}\left(\psi(X_1,\dots,X_n)\right) &= \mathbb{E}\left(\frac{1}{n!}\sum \phi(X_{i_1},\dots,X_{i_n})\right) \\ &= \frac{1}{n!}\sum \mathbb{E}\left(\phi(X_{i_1},\dots,X_{i_n})\right) \\ &= \frac{1}{n!}\sum \mu \\ &= \mu\end{aligned}$$

$$\begin{aligned}\mathrm{Var}\left(\psi(X_1,\dots,X_n)\right) &= \mathbb{E}\left(\frac{1}{n!}\sum \phi(X_{i_1},\dots,X_{i_n})\right)^2 - \mu^2 \\ &= \frac{1}{(n!)^2}\left(\sum \mathbb{E}\left(\phi(X_{i_1},\dots,X_{i_n})^2\right) + \sum \mathbb{E}\left(\phi(X_{i_1},\dots,X_{i_n})\phi(X_{j_1},\dots,X_{j_n})\right)\right) - \mu^2 \\ &= \frac{1}{n!}(\sigma^2+\mu^2) + \sum \mathbb{E}\left(\phi(X_{i_1},\dots,X_{i_n})\phi(X_{j_1},\dots,X_{j_n})\right) - \mu^2 \\ &\leq \frac{1}{n!}(\sigma^2+\mu^2) + \sum \sqrt{\mathbb{E}\left(\phi(X_{i_1},\dots,X_{i_n})^2\right)\mathbb{E}\left(\phi(X_{j_1},\dots,X_{j_n})\right)^2} - \mu^2 \\ &= \frac{1}{n!}(\sigma^2+\mu^2) + \sum \sqrt{(\mu^2+\sigma^2)^2} - \mu^2 \\ &= \sigma^2 + \mu^2 - \mu^2 \\ &= \sigma^2\end{aligned}$$

But since for some value the two variables inside C-S are not the same, the inequality is strict.

Question (1976 STEP IIII Q11)

A population contains individuals of k types, in equal proportions. Among type i, a quantity X is distributed with mean μ_i and variance σ^2 (the same for all i), for $i = 1, 2, ..., k$. It is desired to estimate the mean of X over the whole population. Two methods of estimation are considered. In the first a random sample of size n (with replacement) is drawn from each of the k types, and in the second a random sample of size kn is drawn (with replacement) from the whole population without regard to type. In each case the mean of the kn X-values is computed. Show that the expectation of the resulting estimate is in each case

$$\mu = \frac{1}{k}\sum_{i=1}^{k} \mu_i,$$

but that the second estimate has variance greater than that of the first by an amount

$$\frac{1}{k^2 n}\sum_{i=1}^{k}(\mu_i - \mu)^2.$$

Let $X_{i,j} \sim N(\mu_i, \sigma^2)$ be iid. Let Y_i be a sample from the second distribution,

In the first case:

$$\mathbb{E}\left(\frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} X_{i,j}\right) = \frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n}\mathbb{E}\left(X_{i,j}\right)$$

$$
\begin{aligned}
&= \frac{1}{kn} \sum_{i=1}^{k} \sum_{j=1}^{n} \mu_i \\
&= \frac{1}{kn} \sum_{i=1}^{k} n\mu_i \\
&= \frac{1}{k} \sum_{i=1}^{k} \mu_i
\end{aligned}
$$

In the second case:

$$
\begin{aligned}
\mathbb{E}\left(\frac{1}{kn}\sum_{i=1}^{kn} Y_i\right) &= \frac{1}{kn}\sum_{i=1}^{kn}\mathbb{E}\left(Y_i\right) \\
&= \frac{1}{kn}\sum_{i=1}^{kn}\mathbb{E}\left(Y_i | Y_i \text{ is of type } T\right) \\
&= \frac{1}{kn}\sum_{i=1}^{kn}\mathbb{E}\left(\mu_T\right) \\
&= \frac{1}{kn}\sum_{i=1}^{kn}\frac{1}{k}\left(\sum_{j=1}^{k}\mu_j\right) \\
&= \frac{1}{k}\sum_{j=1}^{k}\mu_j
\end{aligned}
$$

so they are equal.

$$
\begin{aligned}
\text{Var}\left(\frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} X_{i,j}\right) &= \frac{1}{k^2n^2}\sum_{i=1}^{k}\sum_{j=1}^{n}\text{Var}\left(X_{i,j}\right) \\
&= \frac{1}{k^2n^2}\sum_{i=1}^{k}\sum_{j=1}^{n}\sigma^2 \\
&= \frac{\sigma^2}{kn}
\end{aligned}
$$

$$
\begin{aligned}
\text{Var}\left(\frac{1}{kn}\sum_{i=1}^{kn}\sum_{j=1}^{n} Y_i\right) &= \frac{1}{k^2n^2}\sum_{i=1}^{kn}\text{Var}\left(Y_i\right) \\
&= \frac{1}{k^2n^2}\sum_{i=1}^{kn}\mathbb{E}\left(Y_i^2\right) - \frac{1}{kn}\mu^2 \\
&= \frac{1}{k^2n^2}\sum_{i=1}^{kn}\mathbb{E}\left(\mathbb{E}\left(Y_i^2 | Y_i \text{ is of type } T\right)\right) - \frac{1}{kn}\mu^2 \\
&= \frac{1}{kn}\sum_{j=1}^{k}\frac{1}{k}\left(\sigma^2 + \mu_j^2\right) - \frac{1}{kn}\mu^2
\end{aligned}
$$

$$= \frac{1}{kn}\sum_{j=1}^{k}\frac{1}{k}\left(\sigma^2+\mu_j^2\right)-\frac{1}{kn}\mu^2$$
$$= \frac{\sigma^2}{kn}+\frac{1}{k^2n}\sum_{j=1}^{k}\mu_j^2-\frac{1}{kn}\mu^2$$
$$= \frac{\sigma^2}{kn}-\frac{1}{k^2n}\sum_{j=1}^{k}(\mu^2-\mu_j^2)$$

$$\begin{aligned}\sum_{i=1}^{k}(\mu_i-\mu)^2 &= \sum_{i=1}^{k}\mu_i^2-2\sum_{i=1}^{k}\mu_i\mu+k\mu^2\\ &= \sum_{i=1}^{k}\mu_i^2-2k\mu^2+k\mu^2\\ &= \sum_{i=1}^{k}\mu_i^2-k\mu^2\\ &= \sum_{i=1}^{k}\left(\mu_i^2-\mu^2\right)\end{aligned}$$

as required.