If a person masters the fundamentals of his subject and has learned to think and work independently, he will surely find his way and will be better able to adapt himself to progress and changes than the person whose training principally consists in the acquiring of detailed knowledge. –Albert Einstein

Solutions to Problems 21-30

21. Let $\hat{\theta}: S^n \to \Theta$ be an estimator for θ. State the mathematical definitions of its bias $Bias(\hat{\theta})$, variance $\mathbb{V}(\hat{\theta})$, and mean square error $MSE(\hat{\theta})$. Then prove the identity: $MSE(\hat{\theta}) = \mathbb{V}(\hat{\theta}) + Bias(\hat{\theta})^2$. Then construct, respectively, an unbiased estimator for the population mean $\mathbb{E}(X)$ and population variance $\mathbb{V}(X)$.
Solution.

Bias: $Bias(\hat{\theta}) \coloneqq \mathbb{E}(\hat{\theta}) - \theta$ where if θ is regarded as a constant then the bias is a constant too.

Variance: $\mathbb{V}(\hat{\theta}) \coloneqq \mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right)^2$

Mean-Squared-Error: $MSE(\hat{\theta}) \coloneqq \mathbb{E}(\hat{\theta} - \theta)^2$

Proof of Identity. $\mathbb{E}(\hat{\theta} - \mathbb{E}(\hat{\theta}) + \mathbb{E}(\hat{\theta}) - \theta)^2 = \mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right)^2 + 2(\mathbb{E}(\hat{\theta}) - \theta)\mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right) + (\mathbb{E}(\hat{\theta}) - \theta)^2$.

A U-estimator for $\mathbb{E}(X)$: $\hat{\mu} \coloneqq \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$. It is unbiased because we have the familiar result $\mathbb{E}(\hat{\mu}) = \mathbb{E}(X)$.

A U-estimator for $\mathbb{V}(X)$: $\widehat{\sigma^2} \coloneqq \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. It is unbiased because we have the familiar result $\mathbb{E}(\widehat{\sigma^2}) = \mathbb{V}(X)$.

22. Let $\hat{\theta}: S^n \to \Theta$ be an estimator for θ. State the mathematical definition of $\hat{\theta}$ being a consistent estimator. Construct a consistent estimator for the 3rd population moment $\mathbb{E}(X^3)$, is it also unbiased?

Solution. **Consistent estimator**: $\hat{\theta}$ is a consistent estimator for θ iff $\hat{\theta} \xrightarrow{\mathbb{P}} \theta$ as $n \to \infty$.

From WLLN we know that $\frac{1}{n}\sum_{i=1}^{n} X_i^3 \xrightarrow{\mathbb{P}} \mathbb{E}(X^3)$. This is to say that $\frac{1}{n}\sum_{i=1}^{n} X_i^3$ is a consistent estimator for $\mathbb{E}(X^3)$.

Next we check its bias. $\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^3\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^3) = \mathbb{E}(X^3)$. Therefore the sample 3rd moment is an <u>unbiased and consistent</u> estimator for the population 3rd moment.

23. A size-n random sample $\{X_i\}_{i=1}^n$ is obtained from a Bernoulli trial (e.g. by tossing an unfair coin n times or by observing a stable natural yes-no phenomenon n times). The statistician wants to infer the quantity $\theta(1-\theta)$ where $\theta = \mathbb{P}(X=1)$. First, explicitly state S, Θ, and $\mathbb{P}_\theta^X$, then construct an unbiased estimator $U: S^n \to \Theta$ and derive $\mathbb{V}(U)$.

Solution. Population (sample space): $S = \{0,1\}$, the set of vertices of a 1D unit cube. Parameter space: $\Theta = [0,1]$, the unit interval. Population distribution: $d\mathbb{P}_\theta^X = \theta^x(1-\theta)^x dx$. To construct an unbiased estimator for $\theta(1-\theta)$, we first look at $U_1 = \frac{1}{n}\sum_{i=1}^n X_i - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2$. Compute $\mathbb{E}(U_1) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) - \frac{1}{n^2}\left(\sum_{i=1}^n \mathbb{E}(X_i^2) + \sum_{i \neq j} \mathbb{E}(X_i X_j)\right) = \theta - \frac{1}{n^2}(n\theta + n(n-1)\theta^2) = \theta - \frac{1}{n}\theta + n(n-1)\theta^2$, which is not the result we desired. But we can adapt it into $U_2 = \frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n(n-1)}(\sum_{i=1}^n X_i)^2$. Now compute $\mathbb{E}(U_2) = \theta - \frac{\theta}{n-1} - \theta^2$. We make a final adjustment to find

$$U = \frac{1}{n-1}\sum_{i=1}^n X_i - \frac{1}{n(n-1)}\left(\sum_{i=1}^n X_i\right)^2$$

and compute $\mathbb{E}(U) = \theta - \theta^2$.

Derivation of $\mathbb{V}(U) = \mathbb{E}(U^2) - \mathbb{E}(U)^2$. Note that $Y = \sum_{i=1}^n X_i \sim Binom(n, \theta)$, then $U = \frac{1}{n-1}Y - \frac{1}{n(n-1)}Y^2$. $U^2 = \frac{1}{n^2(n-1)^2}(n^2Y^2 - 2nY^3 + Y^4)$.

Lemma. For $Y \sim Binom(n, \theta)$

$\mathbb{E}(Y^2) = n(n-1)\theta^2 + n\theta$.

$\mathbb{E}(Y^3) = n(n-1)(n-2)\theta^3 + 3n(n-1)\theta^2 + n\theta$.

$\mathbb{E}(Y^4) = n(n-1)(n-2)(n-3)\theta^4 + 6n(n-1)(n-2)\theta^3 + 7n(n-1)\theta^2 + n\theta$.

Or compactly : $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ 1 & 7 & 6 & 1 \end{bmatrix} \begin{bmatrix} n\theta \\ n(n-1)\theta^2 \\ n(n-1)(n-2)\theta^3 \\ n(n-1)(n-2)(n-3)\theta^4 \end{bmatrix} = \mathbb{E}\begin{bmatrix} Y \\ Y^2 \\ Y^3 \\ Y^4 \end{bmatrix}$

Proof of Lemma. We use a technique that looks different than MGF but is essentially the same. We denote θ by p and $1-\theta$ by q and allow p and q to vary independently during the derivation and impose the constraint $p+q=1$ *on demand*.

$p_Y(k; n, p, q) = \binom{n}{k} p^k q^{n-k}$.

$\mathbb{E}(Y) = \sum_{k=0}^n \binom{n}{k} k p^k q^{n-k} = p\sum_{k=0}^n \binom{n}{k} k p^{k-1} q^{n-k} = p\sum_{k=0}^n \binom{n}{k} \frac{d}{dp} p^k q^{n-k} = p\frac{d}{dp}\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = p\frac{d}{dp}(p+q)^n = np(p+q)^{n-1} = np$.

$\mathbb{E}(Y^2) = \sum_{k=0}^n \binom{n}{k} k^2 p^k q^{n-k} = \sum_{k=0}^n \binom{n}{k} k(k-1) p^k q^{n-k} + \mathbb{E}(Y) = p^2 \frac{d^2}{dp^2}(p+q)^n + np = n(n-1)p^2(p+q)^{n-2} + np = n(n-1)p^2 + np$.

$$\mathbb{E}(Y^3) = \sum_{k=0}^{n} \binom{n}{k} k^3 p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} k(k-1)(k-2) p^k q^{n-k} + 2\sum_{k=0}^{n} \binom{n}{k} k(k-1) p^k q^{n-k} + \sum_{k=0}^{nb} \binom{n}{k} k^2 p^k q^{n-k} = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np.$$

$$\mathbb{E}(Y^4) = \sum_{k=0}^{n} \binom{n}{k} k^4 p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \{k(k-1)(k-2)(k-3) + 3k(k-1)(k-2) + 2k^2(k-1) + k^3\} p^k q^{n-k} = n(n-1)(n-2)(n-3)p^4 + 6p^3 n(n-1)(n-2) + 7n(n-1)p^2 + np.$$

(end of Proof of Lemma)

Hence, written compactly,

$$\mathbb{E}(U^2) = \frac{1}{n^2(n-1)^2}\begin{bmatrix} n^2 & -2n & 1 \end{bmatrix} \mathbb{E}\begin{bmatrix} Y^2 \\ Y^3 \\ Y^4 \end{bmatrix} = \frac{1}{n^2(n-1)^2}\begin{bmatrix} n^2 & -2n & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ 1 & 7 & 6 & 1 \end{bmatrix}\begin{bmatrix} n\theta \\ n(n-1)\theta^2 \\ n(n-1)(n-2)\theta^3 \\ n(n-1)(n-2)(n-3)\theta^4 \end{bmatrix}$$

Hence, after simplification, $\mathbb{V}(U) = \mathbb{E}(U^2) - [\mathbb{E}(U)]^2 = \frac{\theta(1-\theta)}{n(n-1)}(n(1-2\theta)^2 + 6\theta - 6\theta^2 - 1)$.

24. Same as 23, except that now the estimation target is $\theta(1-\theta)^2$.

Solution. The target this time is not easily adjusted from the trail of 23. We need a universal method. Consider the problem in detail: After the n data are observed, they form 2 groups: one containing all 0 and the other containing all 1. If we consider the sampling-without-replacement experiment: draw three data points from the observed n data and consider what is the probability of having one of them being a 1 and the other two of them being a 0? This probability *seems* to match our estimation target $\theta(1-\theta)^2$. Analytically, this probability is

$$U = \frac{(\sum_{i=1}^{n} X_i)(n - \sum_{i=`}^{n} X_i)(n - \sum_{i=1}^{n} X_i - 1)}{n(n-1)(n-2)}$$

We can verify that its expectation, $\mathbb{E}(U)$, is indeed equal to $\theta(1-\theta)^2$:

Let $Y = \sum_{i=1}^{n} X_i$, then

$$n(n-1)(n-2)\mathbb{E}(U) = \mathbb{E}\big(Y(n-Y)(n-Y-1)\big) = \begin{bmatrix} n^2-n & 1-2n & 1 \end{bmatrix}\mathbb{E}\begin{bmatrix} Y \\ Y^2 \\ Y^3 \end{bmatrix} = \begin{bmatrix} n^2-n & 1-2n & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 3 & 1 \end{bmatrix}\begin{bmatrix} n\theta \\ n(n-1)\theta^2 \\ n(n-1)(n-2)\theta^3 \end{bmatrix}$$

$$= n(n-1)(n-2)\theta(1-\theta)^2$$

Continue with the method of 23 on deriving higher order moments of $Y \sim Binom(n, p, q)$ subject to $p + q = 1$:

$\mathbb{E}(Y^5) = n(n-1)(n-2)(n-3)(n-4)p^5 + 10n(n-1)(n-2)(n-3)p^4 + 25n(n-1)(n-2)p^3 + 15n(n-1)p + np.$

$\mathbb{E}(Y^6) = n(n-1)(n-2)(n-3)(n-4)p^6 + 15n(n-1)(n-2)(n-3)(n-4)p^5 + 65n(n-1)(n-2)(n-3)p^4 + 90n(n-1)(n-2)p^3 + 31n(n-1)p + np.$

We give the result without showing the tedious derivation steps:

$\mathbb{V}(U) = \mathbb{E}(U^2) - \left(\mathbb{E}(U)\right)^2 = \frac{\theta(1-\theta)^2}{n(n-1)(n-2)}\{(9n^2-45n+60)p^3 - (15n^2-69n+84)p^2 + (7n^2-29n+30)p - (n-1)(n-2)\}.$

<u>Advice to students</u>: Any statistician should at least experienced such tediousness once, during their boredom. Any statistician of the computer era should also know how to delegate such manual work to a CAS. Here is an **<u>example using Mathematica (version 6)</u>** to derive the i-th moment of Binomial(n,p):

```
Clear["Global`*"]; MGFbinom[t_]:=(1-p+p Exp[t])^n; i=1; Collect[D[MGFbinom[t],{t,i}]/.t -> 0,p]
```

25. Same Bernoulli trail as in 23, except that now the estimation target is θ and that the statistician is considering using the number of consecutive heads(1s) between two tails (0s), which is a random variable $Z \sim Geometric(\theta)$ with p.m.f $p_Z(k) = (1-\theta)^k\theta, k = 0,1,2, \ldots$, as the basis for constructing an estimator for θ. Is an unbiased estimator possible?

    Solution. Suppose $T(Z)$ is an unbiased estimator for θ. Then

    $$\theta = \sum_{k=0}^{\infty} T(k)\theta^k(1-\theta) \Rightarrow \sum_{k=0}^{\infty} T(k)\theta^k = \frac{\theta}{1-\theta} = \sum_{i=1}^{\infty} \theta^i$$

    Then in order to match the powers, the function $T(k)$ must be defined as equal to 0 when $k = 0$ and equal to 1 when $k \geq 1$. That is, the range of the estimator contains only 2 points. So the answer is "Yes", it is possible to construct an unbiased estimator for θ based on Z, but the estimator will give incredibly senseless estimates for the value of θ.

26. State and prove the fundamental result "Poisson approximation to Binomial" which gives the relationship between the two families of distributions.

Solution. **<u>Poisson approximation to Binomial (PAB)</u>**: Let $X \sim Binomial(n, p)$ and $Y \sim Poisson(np)$ then $X \overset{\mathcal{D}}{\to} Y$ as $n \to \infty$ if $np = \lambda < \infty$ as $n \to \infty$.

<u>Proof of PAB</u>.

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{p^k}{k!} \frac{n!\,(1-p)^{n-k}}{(n-k)!}$$

which tends to $e^{-np} \frac{n^k p^k}{k!}$ because

$$(1-p)^{n-k} = \frac{(1-p)^n}{(1-p)^k} \to (1-p)^n$$
$$\left(1+\frac{x}{n}\right)^n \to e^x \Rightarrow (1-p)^n \to e^{-np}$$
$$n(n-1)\cdots(n-k+1) \to n^k$$

27. Let $X_1, \dots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ population. The sample variance $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Show that its variance $\mathbb{V}(S_n^2) = \frac{2\sigma^4}{n-1}$. *Hint: Chi-sq distribution with degree of freedom n-1.*

Solution. $(n-1)S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2 = \sigma^2 \left[\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2 - n\left(\frac{\bar{X}_n-\mu}{\sigma}\right)^2\right]$.

Let $Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$, then $\frac{n-1}{\sigma^2} S_n^2 = \sum_{i=1}^n Z_i^2 - n\bar{Z}_n^2$. The Cochran's Theorem says that, given a random sample from a normal population, the sample mean and the sample variance are independent. Then

$$\left(\sum_{i=1}^n Z_i^2\right) \perp\!\!\!\perp \bar{Z}_n.$$

Note that $\sum_{i=1}^n Z_i^2 \sim \chi^2(n)$ and $\bar{Z}_n \sim N(0,1)$. Therefore

$$\sum_{i=1}^n Z_i^2 - n\bar{Z}_n^2 \sim \chi^2(n-1)$$

and $\mathbb{V}\left(\sum_{i=1}^n Z_i^2 - n\bar{Z}_n^2\right) = 2(n-1)$. Hence $\mathbb{V}(S_n^2) = \frac{2\sigma^4}{n-1}$.

28. The continuous random variables X, Y, and Z defined in the range $(0 \le X, Y, Z < \infty)$ have the joint probability distribution with the density function $f(X = x, Y = y, Z = z) = (xyz)^{-1/2} \cdot g(x + y + z)$. Derive the marginal distributions of the random variables (i) $U = X + Y + Z$; (ii) $V = Y/X$; and (iii) $W = Z/(X + Y)$.

Solution. Express (X, Y, Z) as functions of (U, V, W): $Z = \frac{UW}{1+W}, X = \frac{U}{(1+W)(1+V)}, Y = \frac{UV}{(1+W)(1+V)}$. Differential linearity gives

$$\left|\det\begin{bmatrix} \frac{1}{(1+w)(1+v)} & -\frac{u}{(1+w)(1+v)^2} & -\frac{u}{(1+w)^2(1+v)} \\ \frac{v}{(1+w)(1+v)} & \frac{u}{(1+w)(1+v)^2} & -\frac{uv}{(1+w)^2(1+v)} \\ \frac{w}{1+w} & 0 & \frac{u}{(1+w)^2} \end{bmatrix}\right| dudvdw = dxdydz$$

$$\det\begin{bmatrix} \frac{1}{(1+w)(1+v)} & -\frac{u}{(1+w)(1+v)^2} & -\frac{u}{(1+w)^2(1+v)} \\ \frac{v}{(1+w)(1+v)} & \frac{u}{(1+w)(1+v)^2} & -\frac{uv}{(1+w)^2(1+v)} \\ \frac{w}{1+w} & 0 & \frac{u}{(1+w)^2} \end{bmatrix} = \frac{1}{(1+w)^3(1+v)^2}\det\begin{bmatrix} 1 & -\frac{u}{1+v} & -\frac{u}{1+w} \\ v & \frac{u}{1+v} & -\frac{uv}{1+w} \\ w & 0 & \frac{u}{1+w} \end{bmatrix}$$

$$= \frac{1}{(1+w)^3(1+v)}\det\begin{bmatrix} 1 & -\frac{u}{1+v} & -\frac{u}{1+w} \\ 0 & \frac{u}{1+v} & 0 \\ w & 0 & \frac{u}{1+w} \end{bmatrix} = \frac{1}{(1+w)^3(1+v)}\frac{u^2}{(1+v)} = \frac{u^2}{(1+v)^2(1+w)^3}$$

Measurability gives $f_{X,Y,Z}(x, y, z)dxdydz = f_{U,V,W}(u, v, w)dudvdw$. Therefore

$$f_{U,V,W}(u, v, w) = g(u)\sqrt{u}\frac{1}{\sqrt{v}(1+v)}\sqrt{\frac{1}{w(1+w)^3}}$$

Therefore U, V, W are mutually independent, and their marginal densities are $f_V(v) = \frac{1}{\pi}\frac{1}{\sqrt{v}(1+v)}\mathbb{I}(v \ge 0)$; $f_W(w) = \frac{1}{2}\sqrt{\frac{1}{w(1+w)^3}}\mathbb{I}(w \ge 0)$; $f_U(u) = 2\pi g(u)\sqrt{u}\mathbb{I}(u \ge 0)$.

29. Let X_1, X_2 be a size-2 random sample from the population density $f_X(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$, for $0 \le x \le \infty$ and θ being a parameter. Derive the sampling distributions of the statistics $U = X_1 + X_2$ and $V = \frac{X_1}{X_1+X_2}$, and hence prove that $U \perp\!\!\!\perp V$. Find the mean and variance of V.

Solution. We solve the problem by density transformation method.
First, express (X_1, X_2) in terms of (U, V): $UV = X_1$, $U - UV = X_2$.
Measurability: $f_{U,V}(u,v)dudv = f_{X_1,X_2}(x_1,x_2)dx_1dx_2$.
Differential linearity: $\left|\det\begin{bmatrix} v & u \\ 1-v & -u \end{bmatrix}\right| dudv = dx_1dx_2 \Rightarrow ududv = dx_1dx_2$.
Therefore $f_{U,V}(u,v) = \frac{u}{\theta^2}e^{-\frac{u}{\theta}}\mathbb{I}(u \ge 0 \text{ and } 0 \le v \le 1)$. Hence $U \perp\!\!\!\perp V$.
Since $\int_0^\infty \frac{u}{\theta^2}e^{-u/\theta}du = 1$, therefore $f_U(u) = \frac{u}{\theta^2}e^{-\frac{u}{\theta}}\mathbb{I}(u \ge 0)$ and $f_V(v) = \mathbb{I}(0 \le v \le 1)$. $\mathbb{E}(V) = \frac{1}{2}$, $\mathbb{V}(V) = \frac{1}{12}$.

30. Let X_1, X_2, X_3 be a size-3 random sample from $N(m, \sigma^2)$. Derive the joint sampling distribution of (i) $U = X_1 - X_3$; (ii) $V = X_2 - X_3$; and (iii) $W = X_1 + X_2 + X_3 - 3m$.

Solution. Let $Y_i = X_i - m \sim N(0, \sigma^2)$. Then

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} U \\ V \\ W \end{bmatrix} \Rightarrow \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} U \\ V \\ W \end{bmatrix} \Rightarrow d\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ -1 & -1 & 1 \end{bmatrix} d\begin{bmatrix} U \\ V \\ W \end{bmatrix}$$

Therefore $\frac{1}{3}dy_1dy_2dy_3 = dudvdw$ and

$$f_{U,V,W}(u,v,w) = \frac{1}{3}f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = \frac{1}{3(2\pi\sigma^2)^{\frac{3}{2}}}e^{-\frac{y_1^2+y_2^2+y_3^2}{2\sigma^2}} = \frac{1}{3(2\pi\sigma^2)^{\frac{3}{2}}}e^{-\frac{2u^2-2uv+2v^2+w^2}{6\sigma^2}}$$

where $y_1^2 + y_2^2 + y_3^2 = \frac{1}{3}(2u^2 - 2uv + 2v^2 + w^2)$ can be derived from expanding $\frac{1}{9}\left\{\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} U \\ V \\ W \end{bmatrix}\right\}^T\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} U \\ V \\ W \end{bmatrix}$. Note that $(U,V) \perp\!\!\!\perp W$.