Big Data 2018: Solutions Exercise Set 2

1. (a) In six ways: (1) 2+2+6 (2) 2+4+4 (3) 2+3+5 (4) 1+4+5 (5) 1+3+6 (6) 3+3+4
   (b) No, you also have to take into account in how many different orders the combinations can be thrown.

2. Taking into account the number of different orders in which the combinations can be thrown, we get:

$$\mathbb{P}(T = 9) = \frac{6+6+3+3+6+1}{216} = \frac{25}{216}$$

$$\mathbb{P}(T = 10) = \frac{3+3+6+6+6+3}{216} = \frac{27}{216}$$

3. We get the following probability distributions:

   (a) The probability distribution of the random variable U is:

u	0	1	2	3	4
$\mathbb{P}(U = u)$	0.01	0.12	0.36	0.32	0.19

   (b) The distribution of V is:

v	0	1	2	3	4
$\mathbb{P}(V = v)$	0.04	0.13	0.44	0.27	0.12

   (c) The distribution of W is:

w	0	1	2	3	4	5	6	7	8
$\mathbb{P}(W = w)$	0.00	0.04	0.06	0.08	0.29	0.16	0.18	0.10	0.09

   (d) The distribution of D is:

d	-4	-3	-2	-1	0	1	2	3	4
$\mathbb{P}(D = d)$	0.00	0.00	0.00	0.13	0.55	0.25	0.07	0.00	0.00

   (e) The conditional distribution of U given $V = 2$ is:

u	0	1	2	3	4
$\mathbb{P}(U = u \mid V = 2)$	0.00	0.09	0.59	0.25	0.07

4. Use Bayes'rule:

$$\mathbb{P}(\text{Rap}|\text{bitch}) = \frac{\mathbb{P}(\text{bitch}|\text{Rap})\mathbb{P}(\text{Rap})}{\mathbb{P}(\text{bitch}|\text{Rap})\mathbb{P}(\text{Rap}) + \mathbb{P}(\text{bitch}|\text{Metal})\mathbb{P}(\text{Metal})}$$
$$= \frac{0.6 \times 0.2}{0.6 \times 0.2 + 0.05 \times 0.8} = 0.75.$$

5. Probability of drawing an apple
(use the law of total probability):

$$\mathbb{P}(A) = \mathbb{P}(A|R)\mathbb{P}(R) + \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|G)\mathbb{P}(G)$$
$$= 0.3 \times 0.2 + 0.5 \times 0.2 + 0.3 \times 0.6 = 0.34$$

Probability of green box given orange (use Bayes' rule):

$$\mathbb{P}(G|O) = \frac{\mathbb{P}(O|G)\mathbb{P}(G)}{\mathbb{P}(O|R)\mathbb{P}(R) + \mathbb{P}(O|B)\mathbb{P}(B) + \mathbb{P}(O|G)\mathbb{P}(G)} = \frac{0.18}{0.36} = 0.5$$

6. (a) For the men we get:

$$\mathbb{P}(\text{Yes}\,|\,\text{Male}) = \frac{\mathbb{P}(\text{Yes, Male})}{\mathbb{P}(\text{Male})} = \frac{0.34}{0.34 + 0.16} = 0.68$$

For the women:

$$\mathbb{P}(\text{Yes}\,|\,\text{Female}) = \frac{\mathbb{P}(\text{Yes, Female})}{\mathbb{P}(\text{Female})} = \frac{0.16}{0.34 + 0.16} = 0.32$$

(b) No, they are not independent. The men have a much higher probability of being accepted than the women.

(c) Yes, we have

$$\mathbb{P}(\text{Yes}\,|\,\text{Male, Program A}) = \mathbb{P}(\text{Yes}\,|\,\text{Female, Program A}) = \mathbb{P}(\text{Yes}\,|\,\text{Program A}) = 0.2$$

(d) Yes, we have

$$\mathbb{P}(\text{Yes}\,|\,\text{Male, Program B}) = \mathbb{P}(\text{Yes}\,|\,\text{Female, Program B}) = \mathbb{P}(\text{Yes}\,|\,\text{Program B}) = 0.8$$

(e) Most women apply to program A, which has a very strict selection. Only 20% of all applicants gets accepted into program A, regardless of gender. Most men apply to program B, which has a very weak selection. A whopping 80% of all applicants gets accepted for program B, regardless of gender. When the probabilities are aggregated over the program dimension the result is a dependence between gender and admission.

7. (a) The population mean and variance are:

$$\mu = \frac{\sum_{i=1}^{N} \mathcal{X}_i}{N} = \frac{4+8+8+8+9+9+9+9}{8} = 8$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (\mathcal{X}_i - \mu)^2}{N} = \frac{(4-8)^2 + 3 \times (8-8)^2 + 4 \times (9-8)^2}{8} = 2.5$$

(b) We get the following probability distribution:

x	4	8	9
$p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{2}$

$$\mathbb{E}(X) = 4 \times \frac{1}{8} + 8 \times \frac{3}{8} + 9 \times \frac{4}{8} = 8$$

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = (4-8)^2 \times \frac{1}{8} + (8-8)^2 \times \frac{3}{8} + (9-8)^2 \times \frac{4}{8} = 2+0+0.5 = 2.5$$

By making a random draw from the population, we obtain a random variable with expected value equal to the population mean and variance equal to the population variance.

(c) We obtain the following distribution of the sample (for convenience we added three columns of sample statistics for later use):

(x_1, x_2)	$p(x_1, x_2)$	$\bar{x}$	$\hat{\sigma}^2$	s^2
$(4, 4)$	$\frac{1}{8} \times \frac{1}{8} = \frac{1}{64}$	4	0	0
$(4, 8)$	$\frac{1}{8} \times \frac{3}{8} = \frac{3}{64}$	6	4	8
$(4, 9)$	$\frac{1}{8} \times \frac{4}{8} = \frac{4}{64}$	6.5	6.25	12.5
$(8, 4)$	$\frac{3}{8} \times \frac{1}{8} = \frac{3}{64}$	6	4	8
$(8, 8)$	$\frac{3}{8} \times \frac{3}{8} = \frac{9}{64}$	8	0	0
$(8, 9)$	$\frac{3}{8} \times \frac{4}{8} = \frac{12}{64}$	8.5	0.25	0.5
$(9, 4)$	$\frac{4}{8} \times \frac{1}{8} = \frac{4}{64}$	6.5	6.25	12.5
$(9, 8)$	$\frac{4}{8} \times \frac{3}{8} = \frac{12}{64}$	8.5	0.25	0.5
$(9, 9)$	$\frac{4}{8} \times \frac{4}{8} = \frac{16}{64}$	9	0	0

(d) We obtain the following sampling distribution for the sample mean:

$\bar{x}$	$p(\bar{x})$
4	$\frac{1}{64}$
6	$\frac{6}{64}$
6.5	$\frac{8}{64}$
8	$\frac{9}{64}$
8.5	$\frac{24}{64}$
9	$\frac{16}{64}$

(e) The expected value of the sample mean is (use the answer to (d)):

$$\mathbb{E}(\bar{X}) = 4 \times \frac{1}{64} + 6 \times \frac{6}{64} + 6.5 \times \frac{8}{64} + 8 \times \frac{9}{64} + 8.5 \times \frac{24}{64} + 9 \times \frac{16}{64} = 8$$

So in this case $\mathbb{E}(\bar{X}) = \mu$. We show later (exercise 10) that this is not a coincidence.

(f) The sampling distribution of $\hat{\sigma}^2$ is:

$\hat{\sigma}^2$	$p(\hat{\sigma}^2)$
0	$\frac{26}{64}$
0.25	$\frac{24}{64}$
4	$\frac{6}{64}$
6.25	$\frac{8}{64}$

(g) The expected value is not equal to σ^2:

$$\mathbb{E}(\hat{\sigma}^2) = 0 \times \frac{26}{64} + 0.25 \times \frac{24}{64} + 4 \times \frac{6}{64} + 6.25 \times \frac{8}{64} = 1.25$$

Hence $\hat{\sigma}^2$ is a biased estimator of σ^2. However, the bias becomes smaller and smaller as the sample size gets bigger: $\hat{\sigma}^2$ is called *asymptotically* unbiased.

(h) The sampling distribution of s^2 is:

s^2	$p(s^2)$
0	$\frac{26}{64}$
0.5	$\frac{24}{64}$
8	$\frac{6}{64}$
12.5	$\frac{8}{64}$

(i) The expected value is equal to σ^2:

$$\mathbb{E}(s^2) = 0 \times \frac{26}{64} + 0.5 \times \frac{24}{64} + 8 \times \frac{6}{64} + 12.5 \times \frac{8}{64} = 2.5$$

As can be shown (proof omitted), s^2 is an unbiased estimator of σ^2 and for that reason often preferred over $\hat{\sigma}^2$.

8. $F = \frac{1}{n}\sum_{i=1}^{n} X_i$. Use the fact that $\mathbb{E}(X_i) = \pi$ and $\text{Var}(X_i) = \pi(1-\pi)$.

(a) $\mathbb{E}(F) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}n\pi = \pi$

(b)

$$\text{Var}(F) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) = \frac{n\pi(1-\pi)}{n^2} = \frac{\pi(1-\pi)}{n}$$

Here we used the rule $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$ and the fact that the X_i are independent.

9.

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}\left[\frac{X-\mu}{\sigma}\right] \\
&= \frac{1}{\sigma}\mathbb{E}[X-\mu] \\
&= \frac{1}{\sigma}(\mathbb{E}[X] - \mathbb{E}(\mu)) \\
&= \frac{1}{\sigma}(\mu - \mu) = 0
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(Z) &= \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) \\
&= \frac{1}{\sigma^2}\mathrm{Var}(X-\mu) \\
&= \frac{1}{\sigma^2}(\mathrm{Var}(X) + \mathrm{Var}(\mu)) \\
&= \frac{1}{\sigma^2}\sigma^2 = 1
\end{aligned}$$

10.

$$\begin{aligned}
\mathbb{E}[\bar{X}] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] \\
&= \frac{1}{n}n\mu = \mu
\end{aligned}$$

$$
\begin{aligned}
\mathrm{Var}[\bar{X}] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] \\
&= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}
$$

11. (a) The expectation is:

$$\mathbb{E}(X) = \int_0^u \frac{1}{u} x \, dx = \left. \frac{x^2}{2u} \right|_0^u = \frac{1}{2}u.$$

(b) We will use the rule that $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. The expectation of X^2 is:

$$\mathbb{E}(X^2) = \int_0^u \frac{1}{u} x^2 \, dx = \left. \frac{x^3}{3u} \right|_0^u = \frac{1}{3}u^2.$$

Hence

$$\mathrm{Var}(X) = \frac{1}{3}u^2 - \frac{1}{4}u^2 = \frac{1}{12}u^2.$$

(c) Call the estimator $\hat{u}$. Clearly $\hat{u}$ cannot be smaller than the sample maximum, since then we would have observations with probability zero in the sample. This would reduce the probability of the whole sample to zero. For every value $\hat{u} \geq \max\{x_1, \ldots, x_n\}$, the probability of the sample is given by

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\hat{u}} = \left(\frac{1}{\hat{u}}\right)^n$$

Clearly, the higher the value of $\hat{u}$, the lower the probability of the sample. Hence we should choose $\hat{u} = \max\{x_1, \ldots, x_n\}$.

(d) The estimate is always below the true value, so it must be downward biased.

(e)

$$
\begin{aligned}
\mathbb{E}(2\bar{X}) &= \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) \\
&= \frac{2}{n}\frac{n}{2}u = u
\end{aligned}
$$

(f)

$$
\begin{aligned}
\mathrm{Var}(2\bar{X}) &= \mathrm{Var}\left(\frac{2}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{4}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{4}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) \\
&= \frac{4}{n^2}\frac{n}{12}u^2 = \frac{u^2}{3n}
\end{aligned}
$$

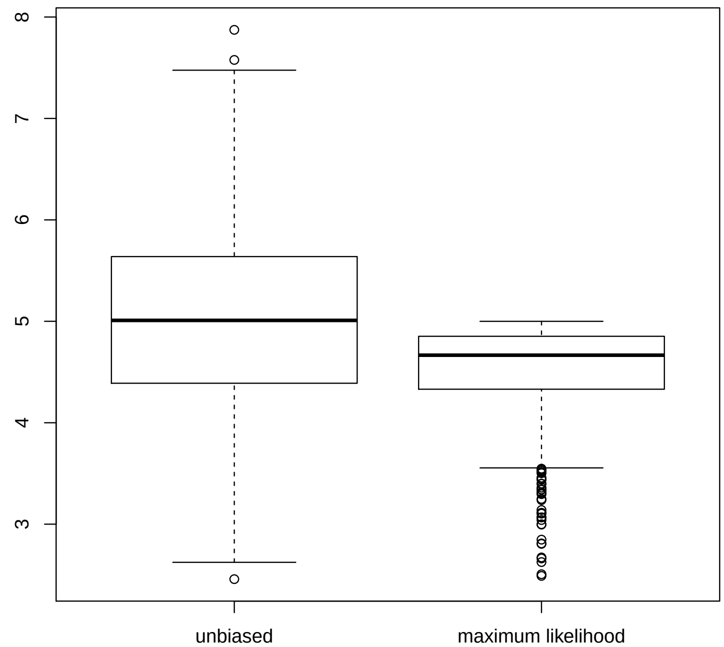
(g) You should find that the mean square error of the maximum likelihood estimator is smaller than that of the unbiased estimator. So many would prefer the ML estimator. Here's a small R program to make the computations:

```
biasvar <- function(m, n, u,seed)
{
# m : number of samples
# n : sample size
# u : max of uniform distribution
# seed: random seed
call <- match.call()
uhat.1 <- uhat.2 <- vector(length = m)
set.seed(seed)
for(i in 1:m) {
# draw sample of size n from uniform distribution U(0,u)
x <- runif(n, min = 0, max = u)
# compute first estimator
uhat.1[i] <- 2 * mean(x)
# compute second estimator
uhat.2[i] <- max(x)
}
m1 <- mean(uhat.1)
b1 <- m1-u
m2 <- mean(uhat.2)
b2 <- m2-u
v1 <- var(uhat.1)
v2 <- var(uhat.2)
mse1 <- b1^2+v1
mse2 <- b2^2+v2
boxplot(uhat.1,uhat.2,names=c("unbiased","maximum likelihood"))
list(call = call, m1 = m1, m2 = m2, v1 = v1, v2 = v2,
     b1 = b1, b2 = b2, mse1 = mse1, mse2 = mse2)
}
```

We added an argument `seed` for reproducibility of results. Here's how you can use the program.

```
> exp.1 <- biasvar(1000,10,5,12345)
> exp.1
$call
biasvar(m = 1000, n = 10, u = 5, seed = 12345)

$m1
[1] 5.006812

$m2
[1] 4.535683

$v1
[1] 0.8573584

$v2
[1] 0.1877325

$b1
[1] 0.006811609

$b2
[1] -0.4643173

$mse1
[1] 0.8574047

$mse2
[1] 0.4033231
```

This call produces the following box plot (we decided to use boxplots rather than histograms).


8
7
6
5
4
3
unbiased
maximum likelihood