STAT 510: Exam 01, Solutions

David Dalpiaz

Monday, March 7

Question 01

Let X_1, X_2, X_3 be **independent** random variables. In particular:

- $X_1 \sim \text{N}(\mu = 2, \sigma = 0.5)$
- $X_2 \sim \text{Poisson}(\lambda = 3)$
- $X_3 \sim \text{Bernoulli}(p = 0.25)$

Calculate each of the following:

(a) [1 point] $P[X_1 > 3]$.

```
pnorm(3, mean = 2, sd = 0.5, lower.tail = FALSE)
```

```
## [1] 0.02275013
```

(b) [1 point] $P[X_2 > 1]$.

```
1 - sum(dpois(0:1, lambda = 3))
```

```
## [1] 0.8008517
```

(c) [1 point] $P[X_3 = 0]$.

```
1 - 0.25
```

```
## [1] 0.75
```

(d) [1 point] $P[X_1 \leq 3, X_2 \leq 4, X_3 = 1] = P[X_1 \leq 3 \cap X_2 \leq 4 \cap X_3 = 1]$.

```
pnorm(3, mean = 2, sd = 0.5) * ppois(4, lambda = 3) * 0.25
```

```
## [1] 0.199179
```

Question 02

Let X and Y be independent random variables such that $X \sim N(0,1)$ and $P[Y=1] = P[Y=-1] = 0.5$. Define $Z = XY$ and note that $Z \sim N(0,1)$.

(a) [1 points] Calculate $\mathrm{Cov}(X,Y)$.

Because X and Y are independent, we know that $\mathrm{Cov}(X,Y) = 0$.

(b) [2 points] Calculate $\mathrm{Cov}(X,Z)$.

$$\begin{aligned}
\mathrm{Cov}(X,Z) &= \mathrm{E}[XZ] - \mathrm{E}[X]\mathrm{E}[Z] \\
&= \mathrm{E}[XZ] \\
&= \mathrm{E}[X^2 Y] \\
&= \mathrm{E}[X^2 Y \mid Y=1] \cdot P[Y=1] + \mathrm{E}[X^2 Y \mid Y=-1] \cdot P[Y=-1] \\
&= 0.5 \cdot \mathrm{E}[X^2] - 0.5 \cdot \mathrm{E}[X^2] \\
&= 0
\end{aligned}$$

(c) [2 points] Are X and Z independent? Justify your answer.

Because $Z = XY$ and $P[Y=1] = P[Y=-1] = 0.5$, we know that

$$P[Z > 0.5 \mid X > 0.5] = 0.5 = P[Z < -0.5 \mid X > 0.5].$$

But we know that $P[Z > 0.5] \neq 0.5$, thus X and Z are dependent. (This is also somewhat directly implied by the definition of Z.)

Question 03

Let W be a random variable with density

$$f(w) = \frac{3w^5}{32}, \quad 0 < w < 2$$

(a) [2 points] Calculate $\mathbb{E}[W]$.

$$\mathrm{E}[W] = \int_{w=0}^{w=2} w \cdot \frac{3w^5}{32} = \frac{12}{7}$$

(b) [3 points] Calculate $\mathbb{V}[W]$.

$$\mathrm{E}\left[W^2\right] = \int_{w=0}^{w=2} w^2 \cdot \frac{3w^5}{32} = 3$$

$$\mathrm{V}[W] = \mathrm{E}\left[W^2\right] - (\mathrm{E}[W])^2 = \frac{3}{49}$$

(c) [3 points] Use Chebyshev's inequality to provide a bound for

$$P\left[|W - \mathbb{E}[W]| \geq \frac{1}{4}\right]$$

$$P\left[|W - \mathbb{E}[W]| \geq \frac{1}{4}\right] \leq \frac{\frac{3}{49}}{\left(\frac{1}{4}\right)^2} = \frac{48}{49}$$

Question 04

Let $X_1, \ldots, X_n$ be a random sample from F and let c be a constant. Define

$$\theta = 2 \cdot P[X > c]$$

Let $\hat{\theta}$ be the plug-in estimator of θ. (Use $\hat{F}$ as notation for the empirical distribution.)

(a) [2 points] Find $\mathbb{E}[\hat{\theta}]$.

$$\hat{\theta} = 2 \cdot (1 - \hat{F}(c))$$

$$\mathbb{E}[\hat{\theta}] = 2 \cdot (1 - F(c))$$

(b) [3 points] Find $\text{SD}[\hat{\theta}]$.

$$\begin{aligned} \text{V}[\hat{\theta}] &= \text{V}[2 \cdot (1 - \hat{F}(c))] \\ &= 4 \cdot \text{V}[\hat{F}(c)] \\ &= 4 \cdot \frac{F(c) \cdot (1 - F(c))}{n} \end{aligned}$$

$$\text{SD}[\hat{\theta}] = 2 \cdot \sqrt{\frac{F(c) \cdot (1 - F(c))}{n}}$$

(c) [3 points] Given $c = 5$, use the limiting distribution of $\hat{\theta}$ to calculate an approximate 95% confidence interval for θ given the data below.

```r
some_data = c(4.9, 3.72, 2.18, 2.54, 2.31, 2.27, 5.14, 4.08, 3.01, 2.89, 5,
              2.37, 5.62, 3.74, 3.96, 4.84, 3.99, 4.62, 4.92, 4.52, 3.92,
              5.89, 2.87, 4.21, 3.16)
```

```r
f_hat_c = mean(some_data <= 5)
theta_hat = 2 * (1 - f_hat_c)
sd_theta_hat = 2 * sqrt(f_hat_c * (1 - f_hat_c) / length(some_data))
```

```r
theta_hat
```

```
## [1] 0.24
```

```r
sd_theta_hat
```

```
## [1] 0.1299846
```

```r
theta_hat + c(-1, 1) * qnorm(0.975) * sd_theta_hat
```

```
## [1] -0.01476516  0.49476516
```

Note that because we are estimating a probability, we would also accept 0 as the lower bound.

Question 05

Consider the coefficient of variation

$$\theta = \frac{\sigma}{\mu}$$

Define an estimator

$$\hat{\theta} = \frac{s}{\bar{x}}$$

where s^2 is the sample variance and $\bar{x}$ is the sample mean. Given the data from the previous question, `some_data`, create 90% bootstrap intervals.

(a) [2 points] Use the "normal" method.

```
boot_rep = function() {
  boot_samp = sample(some_data, replace = TRUE)
  sd(boot_samp) / mean(boot_samp)
}
```

```
theta_hat = sd(some_data) / mean(some_data)
boot_reps = replicate(n = 10000, boot_rep())
se = sd(boot_reps) # we will allow this even though we should be dividing by n
```

```
(norm_int = theta_hat + c(-1, 1) * qnorm(0.95) * se)
```

```
## [1] 0.2341357 0.3385064
```

(b) [2 points] Use the "percentile" method.

```
(perc_int = quantile(boot_reps, c(0.05, 0.95)))
```

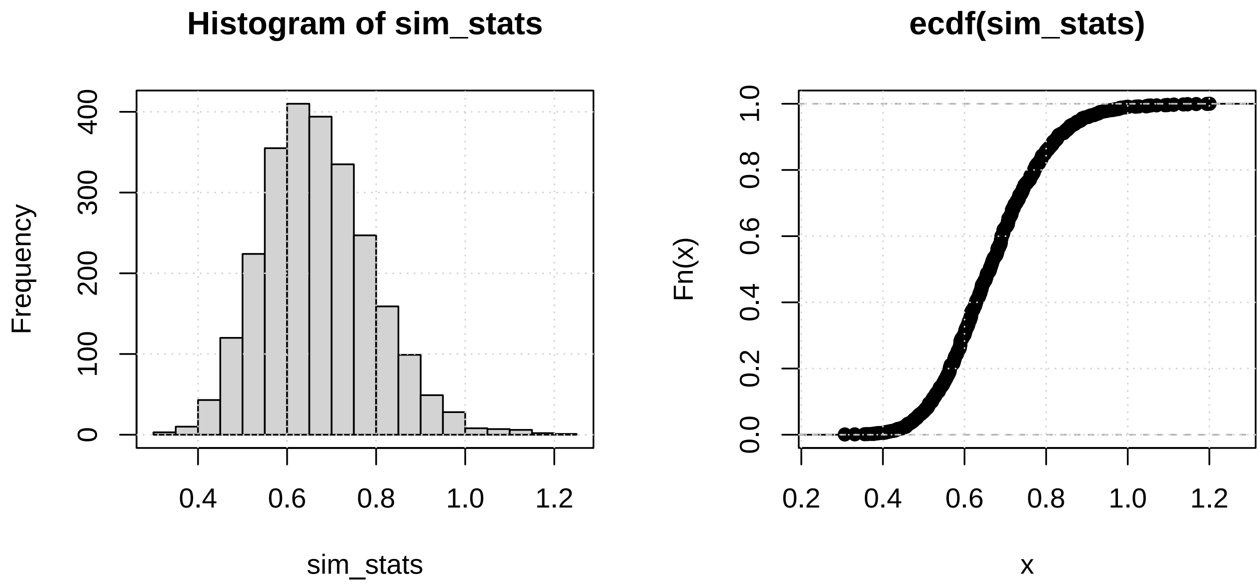
```
##        5%       95%
## 0.2267982 0.3318725
```

(c) [2 points] Now assume that $X_1, X_2, \ldots, X_{20} \sim \text{Pois}(\lambda = 2.2)$. That is, we are consider samples of size $n = 20$. Perform a simulation study that creates an empirical sampling distribution of $\hat{\theta}$. Use at least 2500 simulations. Plot a histogram of the simulated values of the statistic. Also plot the empirical cumulative distribution function. (Place the two plots side-by-side if you have time.)

```
sim_stat = function() {
  sim_data = rpois(n = 20, lambda = 2.2)
  sd(sim_data) / mean(sim_data)
}
```

```
sim_stats = replicate(n = 2500, sim_stat())
```

```
par(mfrow = c(1, 2))
hist(sim_stats)
box()
grid()
plot(ecdf(sim_stats))
grid()
```


Histogram of sim_stats
Frequency
0
100
200
300
400
0.4
0.6
0.8
1.0
1.2
sim_stats
ecdf(sim_stats)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
0.2
0.4
0.6
0.8
1.0
1.2
x

Question 06

Assume $n > 3$. Let $X_1, \ldots, X_n$ be a random sample with mean μ and variance σ^2. Define three estimators for μ:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \dot{X} = \frac{1}{5n}\sum_{i=1}^{n} X_i \qquad \tilde{X} = \frac{2X_1 + 3X_2 + 5X_3}{10}$$

(a) [2 points] Find the **bias**, **variance**, and **MSE** when using $\bar{X}$ to estimate μ.

- $\mathrm{E}[\bar{X}] = \mu$, $\mathrm{bias}[\bar{X}] = 0$
- $\mathrm{V}[\bar{X}] = \sigma^2/n$
- $\mathrm{MSE}[\bar{X}] = \sigma^2/n$

(b) [2 points] Find the **bias**, **variance**, and **MSE** when using $\dot{X}$ to estimate μ.

- $\mathrm{E}[\dot{X}] = \mu/5$, $\mathrm{bias}[\dot{X}] = -4\mu/5$
- $\mathrm{V}[\dot{X}] = \sigma^2/25n$
- $\mathrm{MSE}[\dot{X}] = 16\mu^2/25 + \sigma^2/25n$

(c) [2 points] Find the **bias**, **variance**, and **MSE** when using $\tilde{X}$ to estimate μ.

- $\mathrm{E}[\tilde{X}] = \mu$, $\mathrm{bias}[\tilde{X}] = 0$
- $\mathrm{V}[\tilde{X}] = 19\sigma^2/50$
- $\mathrm{MSE}[\tilde{X}] = 19\sigma^2/50$

(d) [1 points] For what values of μ does $\dot{X}$ outperform $\bar{X}$ in terms of **MSE**.

$$\mathrm{MSE}[\dot{X}] < \mathrm{MSE}[\bar{X}] \implies -\sqrt{\frac{3\sigma^2}{2n}} < \mu < \sqrt{\frac{3\sigma^2}{2n}}$$

(e) [1 points] Is $\bar{X}$ a **consistent** estimator? Justify.

Yes. Clearly $\mathrm{MSE}[\bar{X}] \to 0$, thus $\bar{X}$ is consistent.

(f) [1 points] Is $\dot{X}$ a **consistent** estimator? Justify.

Yes, when $\mu = 0$, otherwise no because the WLLN gives

$$\dot{X} \xrightarrow{p} \frac{\mu}{5}.$$

(g) [1 points] Is $\tilde{X}$ a **consistent** estimator? Justify.

Because the variance is not a function of n, it is clear that $\tilde{X}$ will not converge as needed thus $\tilde{X}$ is not consistent.

(h) [1 points] Done correctly, you will have identified two unbiased estimators, one of which is the sample mean. Does the other unbiased estimator ever outperform the sample mean? Briefly explain.

No. For that to be true we would need

$$\mathrm{MSE}[\tilde{X}] < \mathrm{MSE}[\bar{X}].$$

However, this implies that

$$n < 2.63.$$

This contradicts $n > 3$, therefor it cannot occur.