Exercise Guide for *Statistical Inference (2nd Ed.)* by Casella and Berger

Muthu Chidambaram

Last Updated: August 10, 2019

Contents

About

> *"If you torture the data enough, nature will always confess."*
>
> \- Ronald Coase

There are actual solutions available here. What follows are notes I took as I worked through some of the exercises in the book on my own.

1 Probability Theory

1.1 Exercise 1

(a) Four character strings consisting of H and T (16 total).

(b) $\mathbb{N}$

(c) $\mathbb{R}^+$

(d) $\mathbb{R}^+/0$

(e) $\frac{i}{n}$ for $i = 0, ..., n$.

1.2 Exercise 4

(a) $P(A) + P(B) - P(A \cap B)$

(b) $P(A) + P(B) - 2P(A \cap B)$

(c) Same as (a).

(d) $1 - P(A \cap B)$

1.3 Exercise 6

We have $u + w = 1 \implies u^2 + 2uw + w^2 = 1$. However, it is not possible for $u^2, 2uw, w^2$ to all equal $\frac{1}{3}$; thus, there are no such u, w.

1.4 Exercise 12

(a) Just a case of countable additivity.

(b) Split $\cup_{i=1}^{\infty} A_i$ into $\cup_{i=1}^{n} A_i$ and $\cup_{i=n+1}^{\infty} A_i$. Taking $n \to \infty$ sends the probability of the latter union to 0, leaving the desired result.

1.5 Exercise 13

No, since $P(B^c) = 1 - P(B) \implies P(B) = \frac{3}{4}$ and $\frac{3}{4} + \frac{1}{3} > 1$.

1.6 Exercise 18

This exercise (based on the provided answer) seems to assume that the balls are distinguishable. There are n^n different arrangements of balls in cells (each ball has n options). If exactly one cell must remain empty, then exactly one other cell must have two balls. There are n choices for the empty cell and $n-1$ choices for the double cell. The two balls to go into the double cell can be chosen in $\binom{n}{2}$ ways, and the remaining balls can be ordered in $(n-2)!$ ways (since they are distinguishable), hence giving the provided result.

1.7 Exercise 21

There are $\binom{2n}{2r}$ ways to choose $2r$ shoes from the collection. If we want to ensure that no pairs of shoes are drawn, we should only draw one shoe from each of the n pairs, which we can do in $\binom{n}{2r}$ ways. For each of the $2r$ shoes we can either choose the right or left shoe from the pair (assuming distinguishability).

1.8 Exercise 23

We use Vandermonde's identity with $m = n = r$ to get

$$\sum_{k=0}^{n} \left(\frac{1}{2^n} \binom{n}{k} \right)^2 = \frac{1}{4^n} \binom{2n}{n}$$

1.9 Exercise 25

Depends on how you define the sample space, see Wikipedia for a lengthy discussion.

1.10 Exercise 26

We take the complement of the probability that we roll a 6 in the first 5 rolls: $1 - \sum_{k=0}^{4} \frac{5^k}{6^{k+1}}$.

1.11 Exercise 35

From the definition of conditional probability and the fact that $P(B) > 0$, we have that $P(\cdot|B) > 0$ and $P(S|B) = \frac{P(B)}{P(B)} = 1$. Countable additivity follows from $A \cap B$ and $A' \cap B$ being disjoint if A and A' are.

1.12 Exercise 36

To solve the first part, we take the complement of the probability that the target was hit fewer than two times, which is

$$1 - \left(\frac{4}{5}\right)^{10} - 10\left(\frac{4}{5}\right)^{9} \frac{1}{5} \approx 0.624$$

If we let A be the event that the target was hit at least once, and B be the event that the target was hit at least twice, then we see that $A \cap B = B$. The conditional probability is thus

$$\frac{1 - \left(\frac{4}{5}\right)^{10} - 10\left(\frac{4}{5}\right)^{9} \frac{1}{5}}{1 - \left(\frac{4}{5}\right)^{10}} \approx 0.699$$

2 Transformations and Expectations

2.1 Exercise 1

(a) $F_Y(y) = P(X^3 \leq y) = F_X(y^{\frac{1}{3}})$. Differentiating, we get that $f_Y(y) = 14y(1 - y^{\frac{1}{3}})$ which integrates to 1 over $(0, 1)$.

(b) $f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X(\frac{y-3}{4}) = \frac{7}{4} \exp\left(\frac{-7(y-3)}{4}\right)$, which again integrates to 1 over $(3, \infty)$. The sample space of Y consists of $(3, \infty)$ because $4X+3$ is monotonically increasing and $4(0) + 3 = 3$.

(c) We need only consider positive square roots, since $x \in (0, 1)$. Thus, $f_Y(y) = \frac{1}{2\sqrt{y}} 30y(1 - \sqrt{y})^2$, which integrates to 1 over $(0, 1)$.

2.2 Exercise 2

(a) $f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) = \frac{1}{2\sqrt{y}}$

(b) Plug $x = e^{-y}$ into $f_X(x)$ and multiply by e^{-y} (the negative of the derivative, since $-\log X$ is decreasing).

(c) Plug in $x = \log y$ and multiply by $\frac{1}{y}$.

2.3 Exercise 3

From the definition of Y, we can see that the sample space is $\mathcal{Y} = \left(\frac{n}{n+1}\right)_{\mathbb{N}}$. Solving $y = \frac{x}{x+1}$ for y, we find that $f_Y(y) = f_X(\frac{y}{1-y})$.

2.4 Exercise 4

(a) Integrating from $-\infty$ to 0 gives $\frac{1}{2}$ and likewise for 0 to ∞, so f is a pdf.

(b) If $t \leq 0$, then we have that $P(X < t) = \frac{1}{2}e^{\lambda t}$. Otherwise,

$$
\begin{aligned}
P(X < t) &= \int_{-\infty}^{t} f(x) = \int_{-\infty}^{0} f(x) + \int_{0}^{t} f(x) \\
&= \frac{1}{2} + \frac{1}{2} - \frac{1}{2}e^{-\lambda t} \\
&= 1 - \frac{1}{2}e^{-\lambda t}
\end{aligned}
$$

(c) If $t \leq 0$, $P(|X| < t)$ is clearly 0. Otherwise,

$$
\begin{aligned}
P(|X| < t) &= P(-t < X < t) = P(X < t) - P(X < -t) \\
&= 1 - e^{-\lambda t}
\end{aligned}
$$

from the CDFs computed in part (b).

2.5 Exercise 5

The sample space $\mathcal{Y}$ is $[0, 1]$. Thus, we need to consider $P(Y \leq y) = P(X \leq \sin^{-1}\sqrt{y})$ for $y \in [0, 1]$. By symmetry, we can just consider the case where $\sin^{-1}$ is restricted to $[0, \frac{\pi}{2}]$, as the other three quadrants have the same area. Therefore,

$$\begin{aligned} f_Y(y) &= 4\frac{\mathrm{d}}{\mathrm{d}y}F_X(\sin^{-1}\sqrt{y}) \\ &= 4\frac{\mathrm{d}}{\mathrm{d}y}\frac{\sin^{-1}\sqrt{y}}{2\pi} \\ &= \frac{1}{\pi\sqrt{y(1-y)}} \end{aligned}$$

At $y = 0$ and $y = 1$, the density is infinite/undefined - I'm not really sure how to interpret this.

2.6 Exercise 11

(a) We have that

$$\begin{aligned} \mathrm{E}[X^2] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} &= \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^2 e^{-\frac{x^2}{2}} \\ &= \frac{2}{\sqrt{2\pi}}\big(\lim_{x\to\infty} -xe^{-\frac{x^2}{2}} + \int_0^{\infty} e^{-\frac{x^2}{2}}\big) \\ &= 1 \end{aligned}$$

Where we integrated by parts using $dv = xe^{-\frac{x^2}{2}}$ and $u = x$, applied L'Hôpital's rule, and then compared to the CDF of the normal distribution.

(b) The pdf $f_Y(y)$ is just $f_X(y) + f_X(-y)$, so $f_Y(y) = \frac{2}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$. We can then compute $\mathrm{E}[Y]$ as

$$\mathrm{E}[Y] = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} ye^{-\frac{y^2}{2}} = \frac{2}{\sqrt{2\pi}}$$

To compute the variance we need to compute $\mathrm{E}[Y^2]$, which is identical to $\mathrm{E}[X^2]$. Thus, $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 = 1 - \frac{2}{\pi}$.

2.7 Exercise 12

We see that $y = d\tan(x)$. Since $\tan(x)$ is increasing on $(0, \frac{\pi}{2})$, we get that

$$\begin{aligned} f_Y(y) &= F_X\big(\tan^{-1}\frac{y}{d}\big)\frac{\mathrm{d}}{\mathrm{d}y}\tan^{-1}\frac{y}{d} \\ &= \frac{2\tan^{-1}\frac{y}{d}}{\pi d\big(1+\frac{y^2}{d^2}\big)} \end{aligned}$$

for $y \in (0, \infty)$. We can compute $\mathrm{E}[Y]$ directly from $f_X(x)$ as

$$\begin{aligned} \mathrm{E}[Y] &= \frac{2d}{\pi} \int_0^{\frac{\pi}{2}} \tan(x) dx \\ &= \infty \end{aligned}$$

So $\mathrm{E}[Y]$ does not exist.

2.8 Exercise 13

A sequence of flips has length l if there are either l heads in a row or l tails in a row, so $P(X = l) = p^l(1-p) + (1-p)^l p = p\mathrm{Geom}(1-p) + (1-p)\mathrm{Geom}(p)$. Therefore, by linearity of expected value, $\mathrm{E}[X] = \frac{p}{1-p} + \frac{1-p}{p}$.

2.9 Exercise 14

(a) We have that

$$\int_0^\infty 1 - F_X(x) dx = \int_{x=0}^\infty \int_{y=x}^\infty f_X(y) dy dx$$

Since $0 \leq x \leq \infty$ and $x \leq y \leq \infty$, we can change the order of integration by having $0 \leq y \leq \infty$ on the outside and $0 \leq x \leq y$ on the inside

$$\begin{aligned} \int_{x=0}^\infty \int_{y=x}^\infty f_X(y) dy dx &= \int_{y=0}^\infty \int_{x=0}^y dx f_X(y) dy \\ &= \int_0^\infty y f_X(y) dy \\ &= E[X] \end{aligned}$$

(b) We can rewrite the expected value as a sum of infinite sums to see that

$$\begin{aligned} E[X] &= \sum_{k=1}^\infty k f(k) = \sum_{k=1}^\infty f(k) + \sum_{k=2}^\infty f(k) + \ldots \\ &= \sum_{k=0}^\infty \sum_{j=k+1}^\infty f(k) = \sum_{k=0}^\infty (1 - F_X(k)) \end{aligned}$$

2.10 Exercise 17

(a) We need to solve $m^3 = 1 - m^3$, since $F_X(x) = x^3$. Solving gives $m = \sqrt[3]{\frac{1}{2}}$.

(b) f is an even function, so $m = 0$.

2.11 Exercise 18

We differentiate under the integral sign to get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}a}\mathrm{E}[|X-a|] &= \frac{\mathrm{d}}{\mathrm{d}a}\left(\int_a^\infty (x-a)f_X(x)dx + \int_{-\infty}^a (a-x)f_X(x)dx\right) \\ &= \int_a^\infty -f_X(x)dx + \int_{-\infty}^a f_X(x)dx \\ &= 2F_X(a) - 1\end{aligned}$$

Setting the last line equal to 0 yields $a = m$. To see that this is a minimum, we note that $2F_X(a) - 1 \leq 0$ when $a < m$ and $2F_X(a) - 1 \geq 0$ when $a > m$.

2.12 Exercise 25

(a)

$$\begin{aligned}F_{-X}(x) = P(X \geq -x) &= 1 - F_X(-x) \\ &= \int_{-x}^\infty f_X(x)dx \\ &= \int_{-x}^0 f_X(x)dx + \int_0^\infty f_X(x)dx \\ &= \int_0^x f_X(x)dx + \int_{-\infty}^0 f_X(x)dx \\ &= F_X(x)\end{aligned}$$

(b)

$$\begin{aligned}M_X(t) &= \int_{-\infty}^\infty e^{tx} f_X(x)dx \\ &= \int_{-\infty}^0 e^{tx} f_X(x)dx + \int_0^\infty e^{tx} f_X(x)dx \\ &= \int_0^\infty e^{-tx} f_X(x)dx + \int_{-\infty}^0 e^{-tx} f_X(x)dx \\ &= M_X(-t)\end{aligned}$$

2.13 Exercise 29

(a) For binomial, we have

$$\begin{aligned}
\mathrm{E}[X(X-1)] &= \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^(k-2)(1-p)^{(n-2)-(k-2)} \\
&= n(n-1)p^2
\end{aligned}$$

For Poisson, we have

$$\begin{aligned}
\mathrm{E}[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!} \\
&= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2} e^{-\lambda}}{(k-2)!} \\
&= \lambda^2
\end{aligned}$$

(b) To compute the variances, we can use linearity of expectation to see that $\mathrm{E}[X(X-1)] + \mathrm{E}[X] - \mathrm{E}[X]^2 = \mathrm{Var}[X]$. So for binomial, we get

$$\mathrm{Var}[X] = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p)$$

And for Poisson, we get

$$\mathrm{Var}[X] = \lambda^2 + \lambda - \lambda^2 = \lambda$$

(c) This one is kind of a pain to typeset; the technique is the same as (a), although more involved.

2.14 Exercise 31

No such distribution exists, since $M_X(0) = 0$, which contradicts the fact that $\mathrm{E}[1] = 1$.

3 Common Families of Distributions

3.1 Exercise 1

$$\mathrm{E}[X] = \sum_{i=N_0}^{N_1} \frac{i}{N_1 - N_0 + 1} = \frac{1}{N_1 - N_0 + 1}(\sum_{i=1}^{N_1} i - \sum_{i=1}^{N_0-1} i) = \frac{N_1 + N_0}{2}$$

$$\mathrm{E}[X^2] = \frac{1}{N_1 - N_0 + 1}(\sum_{i=1}^{N_1} i^2 - \sum_{i=1}^{N_0-1} i^2) = \frac{2N_0^2 + 2N_0N_1 - N_0 + 2N_1^2 + N_1}{6}$$

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{(N_1 - N_0 + 1)^2 - 1}{12}$$

3.2 Exercise 2

(a) Suppose the number of defective parts in the lot is D. The likelihood that we accept a sample from such a lot is $P(X = 0)$, where X is distributed according to Hypergeom$(100, D, K)$. Given that a lot is unacceptable if $D > 5$, we need only consider the case where $D = 6$, since larger D lead to lower likelihood of being incorrectly accepted. Thus, we solve

$$\frac{\binom{94}{K}}{\binom{100}{K}} < 0.1$$

by running the following Python code:

```python
from scipy import special

K = 1
while special.binom(94, K) / special.binom(100, K) > 0.1:
  K += 1
```

to get that $K = 32$.

(b) This is the same as part (a) except that we now have to consider $P(X = 0) + P(X = 1) < 0.1$. Slightly modifying the Python code above yields $K = 51$.

3.3 Exercise 3

At each second, the probability that at least one car passes in the next 3 seconds is $1 - P(X = 0)$, where X is binomial with $n = 3$ and $p = p$. Thus, the probability that the pedestrian has to wait 4 seconds before starting to cross is the probability that at least one car passes in the first three seconds multiplied by the probabilities that a car crosses at the 4th second and that no cars cross during seconds 5-7. This probability is $(1 - (1 - p)^3)p(1 - p)^3$.

NOTE: I compared my answer to the actual solutions, but I don't buy that the actual answer is $(1 - p(1 - p)^3)(1 - p)^3$ based on the logic provided. What if cars pass on the first and second seconds, but then no cars pass after? The pedestrian could then leave after two seconds.

3.4 Exercise 4

(a) If unsuccessful keys are not eliminated, then this is the same as asking for the expectation of a geometric random variable with $p = \frac{1}{n}$, so we expect that it will take n trials.

(b) If we eliminate unsuccessful keys, then our expectation looks like

$$\begin{aligned}\mathrm{E}[X] &= \frac{1}{n} + 2 * \frac{n-1}{n} * \frac{1}{n-1} + ... \\ &= \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2}\end{aligned}$$

3.5 Exercise 5

The standard drug's effectiveness can be modeled as a binomial distribution with $n = 100$ and $p = 0.8$, since we are given that the drug is expected to be effective in 80 out of 100 cases. Computing $P(X \geq 85)$ yields a value of approximately 0.129, which means that it is fairly unlikely for the new drug to work in 85 (or more) cases if it was worse than the first drug. Thus, we conclude that the new drug is superior.

3.6 Exercise 7

We would like to have $P(X \geq 2) = 1 - P(X < 2) > 0.99$. Thus, we can solve $e^{-\lambda}(1 + \lambda) < 0.01$ to get that $\lambda > 6.638$.

3.7 Exercise 12

I tried to do this analytically but struggled to show equivalence between the summations; obtaining the result through qualitative reasoning is much simpler. $F_X(r-1)$ is the probability that there are at most $r-1$ successes in n Bernoulli trials, which is the same as saying that we get success number r after n trials. Thus, we must have at least $n+1-r$ failures, which is $1 - F_Y(n-r)$ (I don't deserve credit for this, I looked up part of the solution here - maybe I'll retry the analytic approach later).

3.8 Exercise 15

I spent an absurd amount of time on this question because I was working off of Wikipedia's provided negative binomial MGF, which treats p as $1-p$. After finding the proper MGF, the result is straightforward:

$$\begin{aligned}
\lim_{r\to\infty}\left(\frac{p}{1-(1-p)e^t}\right)^r &= \lim_{r\to\infty}\left(1+\frac{p}{1-(1-p)e^t}-\frac{1-(1-p)e^t}{1-(1-p)e^t}\right)^r \\
&= \lim_{r\to\infty}\left(1+\frac{r\frac{-(1-p)+(1-p)e^t}{1-(1-p)e^t}}{r}\right)^r \\
&= \exp\bigl(-\lambda+\lambda e^t\bigr)
\end{aligned}$$

3.9 Exercise 19

Since $\alpha = 1, 2, 3, ...$ we have $\Gamma(\alpha) = (\alpha-1)!$. We can then integrate by parts with $u = \frac{1}{(\alpha-1)!}z^{\alpha-1}$ and $dv = e^{-z}$ to get

$$\begin{aligned}
\int_x^\infty \frac{1}{(\alpha-1)!}z^{\alpha-1}e^{-z}dz &= \frac{x^{\alpha-1}e^{-x}}{(\alpha-1)!}+\int_x^\infty \frac{1}{(\alpha-2)!}z^{\alpha-2}e^{-z}dz \\
&= \sum_{y=0}^{\alpha-1}\frac{x^y e^{-x}}{y!}
\end{aligned}$$

where the last line follows from repeatedly integrating by parts until $z^{\alpha-k} = 0$. Probabilistically, this equality gives that $P(X \ge x) = P(Y \le \alpha - 1)$ where X is Gamma distributed with $\alpha = \alpha, \beta = 1$ and Y is Poisson with $\lambda = x$.

3.10 Exercise 44

Directly applying Chebychev to $P(|X| \ge b)$ gives $P(|X| \ge b) \le \frac{\mathrm{E}[|X|]}{b}$. Furthermore, since $P(|X| \ge b) = P(X^2 \ge b^2)$, we can apply Chebychev again to get $P(|X| \ge b) \le \frac{\mathrm{E}[X^2]}{b^2}$. Both $\mathrm{E}[|X|]$ and $\mathrm{E}[X^2]$ can be computed via integration by parts as 1 and 2 respectively. Thus, for $b = 3$ the X^2 bound is better and for $b = \sqrt{2}$ the $|X|$ bound is better.

3.11 Exercise 45

(a)

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^\infty e^{tx}f_X(x)dx \\
&\ge e^{at}\int_{x\ge a} f_X(x)dx \quad \text{since } e^{xt} \ge e^{at} \text{ for } x \ge a, t \ge 0 \\
&\ge e^{at}P(X \ge a)
\end{aligned}$$

(b) Identical to (a), except we use $x \le a$ in step 2 since $t < 0$.

(c) We just need $h(t, x)$ to be nonnegative and $h(t, x) \ge 1$ for all $x, t \ge 0$, since then we can replicate the steps of (a).