Prof. David Draper
Department of
Applied Mathematics and Statistics
University of California, Santa Cruz

AMS 132: Discussion Section 1 (with correction in part (f))

1. Consider a random variable Y that represents your uncertainty about a *waiting time*: a measurement of how long you'll have to wait (starting at time 0) until something interesting happens (e.g., until a particular Amazon customer makes a new purchase, or until a particular type of decay of a radioactive atom occurs, or ...). Under assumptions that reasonably describe some (but not all) real-world waiting-time phenomena (see the AMS 131 and 132 textbook DeGroot and Schervish (2012), pages 321–325, for details), Y can be modeled as a draw from the *exponential distribution*: for any $\beta > 0$, the probability density function (PDF, or just *density*, or sometimes just *distribution* for short) for Y is

$$p_Y(y \mid \beta) \triangleq p(y \mid \beta) = \left\{ \begin{array}{cc} \beta\, e^{-\beta\, y} & y \geq 0 \\ 0 & \text{otherwise} \end{array} \right\} = \beta\, e^{-\beta\, y}\, I(y \geq 0)\,, \qquad (1)$$

in which $I(A)$ is the indicator function, equal to 1 if A is true and 0 otherwise, and $\triangleq$ means "is defined to be" (for simplicity, we'll often drop the subscript Y in $p_Y(y \mid \beta)$ when this creates no confusion). Recalling the ideas in AMS 131, you can see that Y is a continuous random variable on the non-negative part ($y \geq 0$) of the real number line and that $\{p(y \mid \beta), \beta \geq 0\}$ is actually a *family* of densities indexed by the *parameter* β. We'll often use the succinct notation

$$(Y \mid \beta) \sim \text{Exponential}(\beta) \qquad (2)$$

to mean the same thing as equation (1); here, as in AMS 131, $\sim$ stands for "is distributed as," so that equation (2) is read as Y *(given* β*) is Exponentially distributed with parameter* β or (even more succinctly) as Y *is Exponential*(β).

(a) Sketch the density function for Y with $\beta = 1$ and $\beta = 2$. What role does β appear to play in this family of distributions?

(b) It can be shown, by repeated integration by parts, that for any non-negative integer k,

$$E\big(Y^k\big) = \int_0^\infty y^k \left[\beta\, e^{-\beta\, y}\right] dy = \frac{k!}{\beta^k} \qquad (3)$$

(it might be good for you to work this out with $k = 1$, just for practice). Use equation (2) to show that

$$E(Y) \triangleq \theta = \frac{1}{\beta}\,, \quad V(Y) \triangleq \sigma^2 = \frac{1}{\beta^2} \quad \text{and} \quad SD(Y) \triangleq \sigma = \frac{1}{\beta}\,. \qquad (4)$$

The fact that $E(Y) = \frac{1}{\beta}$ means that β represents the *rate* at which the waiting time process unfolds: since Y is measured in time units (e.g., seconds), the units of β must be $\frac{1}{\text{time}}$ (e.g., 2.6 interesting events per second).

Suppose now that you're about to collect an IID random sample $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ of n observations from this waiting time process (for n a non-negative integer), and let $\boldsymbol{y} = (y_1, \ldots, y_n)$ stand for the observed data values. Then your model for $\boldsymbol{Y}$ can be written

$$p_{Y_i}(y_i \,|\, \beta) \triangleq p(y_i \,|\, \beta) = \left\{ \begin{array}{cc} \beta\, e^{-\beta\, y_i} & y_i \geq 0 \\ 0 & \text{otherwise} \end{array} \right\} \quad (\text{for } i = 1, \ldots, n)\,, \tag{5}$$

and this can be summarized even more succinctly as

$$(Y_i \,|\, \beta) \stackrel{\text{IID}}{\sim} \text{Exponential}(\beta) \quad (\text{for } i = 1, \ldots, n)\,. \tag{6}$$

The point of gathering this data set is to use $\boldsymbol{y}$ to draw valid statistical inferences about β. In the frequentist approach, as usual this requires you to first think probabilistically, temporarily pretending that β is known and considering the possible data sets $\boldsymbol{Y}$ you could observe. We'll see later in the course that a good frequentist estimate for β based on $\boldsymbol{Y}$ is

$$\hat{\beta} = \frac{n}{S} = \overline{\boldsymbol{Y}}^{-1}\,, \tag{7}$$

in which $S = \sum_{i=1}^n Y_i$ is the sample sum and $\overline{\boldsymbol{Y}} = \frac{1}{n} \sum_{i=1}^n Y_i$ is the sample mean.

(c) Recall from AMS 131 that — as long as $E(Y_i)$ exists and is finite, which is certainly true here — the expected value $E\left(\overline{\boldsymbol{Y}}\right)$ of the sample mean is the same as the expected value $E(Y_i)$ of any single observation Y_i going into the calculation of the mean:

$$E\left(\overline{\boldsymbol{Y}}\right) = E(Y_i) = \theta\,. \tag{8}$$

Given your result for the expected value of Y in part (b) above, briefly explain why $\overline{\boldsymbol{Y}}^{-1}$ is an intuitively reasonable estimator of β.

(d) Also recall from AMS 131 that — as long as $V(Y_i)$ exists and is finite, which is certainly also true here — the variance $V\left(\overline{\boldsymbol{Y}}\right)$ of the sample mean is related to the variance of any single observation Y_i going into the calculation of the mean through the expression

$$V\left(\overline{\boldsymbol{Y}}\right) = \frac{V(Y_i)}{n} = \frac{\sigma^2}{n} = \frac{1}{\beta^2\, n}\,. \tag{9}$$

Does this match your intuition that, as the amount of information you have about β — which is driven by the sample size n — goes up, your uncertainty about β on the basis of $\overline{\boldsymbol{Y}}$ should go down?

(e) For large n, what should the repeated-sampling distribution of $\overline{\boldsymbol{Y}}$ look like? Explain briefly.

(f) The estimator $\hat{\beta}$ in (7) is related to $\overline{\boldsymbol{Y}}$ through the invertible transformation

$$\hat{\beta} = h\left(\overline{\boldsymbol{Y}}\right) \quad \text{for} \quad h(t) = \frac{1}{t}\,. \tag{10}$$

Use the *Delta Method* (DeGroot and Schervish (2012), pages 364–365) to show that for large n,

$$E\left(\hat{\beta}\right) \doteq \beta \quad \text{and} \quad V\left(\hat{\beta}\right) \doteq \frac{\beta^2}{n}\,. \tag{11}$$

The approximate expression for the repeated-sampling variance of $\hat{\beta}$ in (11) gets larger as β increases. Does this make good intuitive sense? Explain briefly.