

Announcements

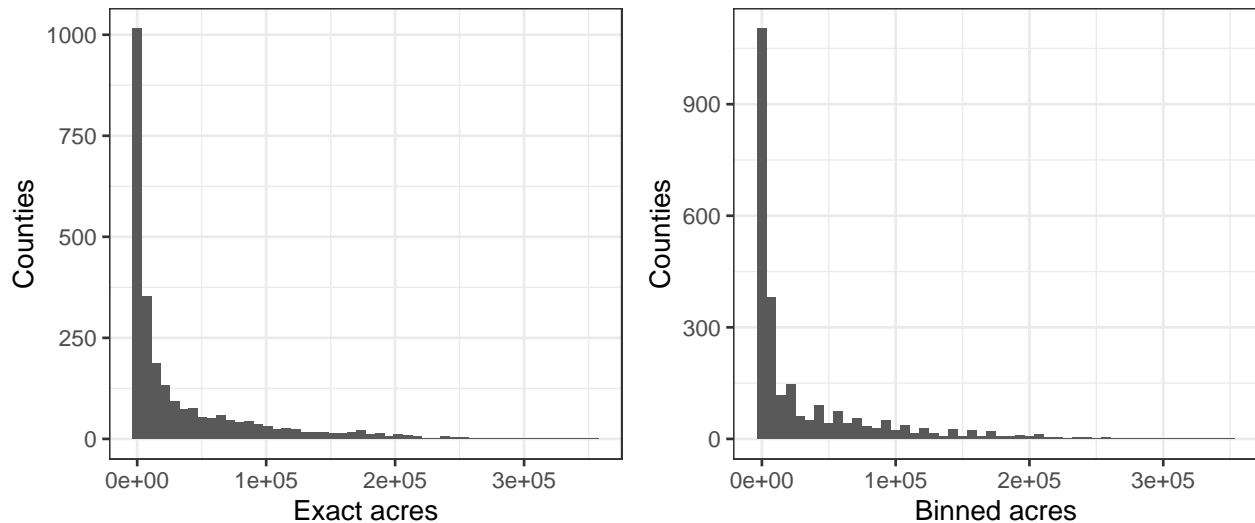
Make sure to sign in on the [google form](#).

Pset 5 due...



It's Corn - A Big Lump with Knobs

The following questions deal with data on how many acres of corn are planted in each U.S. county [available here](#).



Continuous data are often binned into intervals for privacy or precision reasons. In this set of problems, we'll be modeling corn acreage for U.S. counties under various assumptions.

1. First, assume there has been no binning. Therefore, the acreage devoted to corn farming in a county is i.i.d. $Y_i \sim \text{Expo}(\lambda)$. Show that this model is a Natural Exponential Family by finding θ , $\Psi(\theta)$, and $h(y)$ such that $f_{Y_i}(y) = e^{\theta y - \Psi(\theta)} h(y)$ where $h(y)$ is a PDF or PMF.

- Page 2

6. As the bin width decreases ($\tau \rightarrow 0$), the likelihood function should converge to the likelihood function of i.i.d. exponentials:

$$\lambda^n \exp(-\lambda n \bar{y})$$

By Taylor expanding the likelihood from (4) with respect to τ , show this is the case.

7. Do the X_i follow a Natural Exponential Family? If so, use $\frac{1}{2^{x+1}}$ as $h(x)$.

8. Find the mean and variance of X_i .

9. Find the MLE for λ .

10. What is the MSE of $\hat{\lambda}_{\text{MLE}}$?

11. Would Rao-Blackwellization improve the MLE for λ ? If not, suggest your own improvement.

12. Using the real data, how does the MLE compare to (1) using the points as they are to estimate λ and (2) using the midpoint of each interval to estimate λ ?

##	True lambda	MLE	Bin left	Bin center
## Value	2.9e-05	0.000029	0.000031	0.000029
## Relative difference from true lambda	0.0e+00	-0.015354	0.058879	-0.017043

Geometric Underpinnings

It turns out that the distribution above in (7) is just a special form of the Geometric with a support restricted to multiples of τ . In this question, we'll explore the Geometric further. Let Y_1, \dots, Y_n be i.i.d. $\text{Geom}(p)$ with p unknown.

1. Show that the model is a Natural Exponential Family by finding θ , $\Psi(\theta)$, and $h(y)$.
2. Use results about NEFs to derive the mean and variance.
3. It follows from (1) that \bar{Y} is a sufficient statistic. Check this directly using the factorization criterion.
4. Consider the estimator of the mean (estimating q/p) Y_1 . Use Rao-Blackwellization to improve this estimator.

5. Consider the estimator of the mean (estimating q/p) $\vec{w} \cdot \vec{Y}$ where \vec{w} is a vector of weights that sums to 1. Use Rao-Blackwellization to show that the best \vec{w} is $(1/n, \dots, 1/n)$.