Alex Tsun
CSE 312: Foundations of Computing II

Section #8
August 13th, 2020

Section #8 Solutions

Shreya Jayaraman, Luxi Wang, Alex Tsun

1. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be iid samples from $Exp(\Theta)$ where Θ is a random variable (not fixed).

   (a) Using the prior $\Theta \sim Gamma(r, \lambda)$ (for some arbitrary but known parameters $r, \lambda > 0$), show that the posterior distribution $\Theta|\mathbf{x}$ also follows a Gamma distribution and identify its parameters (by computing $\pi_\Theta(\theta|\mathbf{x})$). Then, explain this sentence: "The Gamma distribution is the conjugate prior for the rate parameter of the Exponential distribution". Hint: This can be done in just a few lines!

   (b) Now derive the MAP estimate for Θ. The mode of a $Gamma(s, \nu)$ distribution is $\frac{s-1}{\nu}$. Hint: This should be just one line using your answer to part (a).

   (c) Explain how this MAP estimate differs from the MLE estimate (recall for the Exponential distribution it was just the inverse sample mean $\frac{n}{\sum_{i=1}^n x_i}$, and provide an interpretation of r and λ as to how they affect the estimate.

Solution:

(a) Remember the posterior is proportional to likelihood times prior:

$$\begin{aligned}
\pi_\Theta(\theta|x) &\propto L(x|\theta)\pi_\Theta(\theta) && \text{[def of posterior]} \\
&= \left(\prod_{i=1}^n \theta e^{-\theta x_i}\right) \cdot \frac{\lambda^r}{(r-1)!}\theta^{r-1}e^{-\lambda\theta} && \text{[def of likelihood, Gamma pdf]} \\
&\propto \theta^n e^{-\theta \sum x_i} \theta^{r-1} e^{-\lambda\theta} && \text{[algebra]} \\
&= \theta^{(n+r)-1} e^{-(\lambda + \sum x_i)\theta}
\end{aligned}$$

Therefore $\Theta|\mathbf{x} \sim Gamma(n + r, \lambda + \sum x_i)$, since the final line above is proportional to the pdf for the gamma distribtution.

(b) Just citing the mode of a Gamma, we get

$$\frac{n + r - 1}{\sum x_i + \lambda}$$

(c) We see how the estimate changes from the MLE: pretend we saw $r - 1$ extra events over λ units of time. (Instead of waiting for n events, we waited for $n + r - 1$, and instead of $\sum x_i$ as our total time, we now have $\lambda + \sum x_i$ units of time).

2. Suppose the true fraction of the U.S. population who like broccoli is p. We wish to estimate p by taking a random sample: we randomly choose n individuals, ask each of them whether they like broccoli or not, and estimate p with the maximum-likelihood estimator $\hat{p}$ (the number of people who liked broccoli divided by n). We would like to quantify the accuracy of this estimate.

   Suppose we interviewed 100 people and 20 of them liked broccoli ($\hat{p} = 0.2$). Find an 86% confidence interval for the true p.

   Solution:
   With $n = 100$ and a desired confidence level of 86%, we seek Δ such that:

$$P(|\hat{p} - p| < \Delta) = P(p \in< [\hat{p} - \Delta, \hat{p} + \Delta]) = 0.86$$

   We know that $E[\hat{p}] = p$ (our point estimation).

   By the CLT, we have $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is approximately normally distributed. We will use the formula for a confidence interval

$$\left[\hat{\theta} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \hat{\theta} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

   Since $100(1 - \alpha) = 86$, we have $\alpha = 0.14$ and $1 - \alpha/2 = 0.93$. We don't know p, so we'll use our estimate $\hat{p}$ instead to estimate the standard deviation (of $Ber(p)$): $\hat{\sigma} = \sqrt{\hat{p}(1 - \hat{p})} = \sqrt{0.2 \cdot 0.8} = \sqrt{0.16} = 0.4$.

$$\hat{p} = 0.2, \qquad \Delta = \Phi^{-1}(0.93)\frac{\hat{\sigma}}{\sqrt{n}} = 1.475\frac{0.4}{\sqrt{100}} = 0.059$$

   Hence our 86% confidence interval is

$$[0.2 - 0.059, 0.2 + 0.059] = [0.141, 0.259]$$

3. Suppose you live on a tree farm with a large field. You've always used Fertilizer Y, but your friend recently recommended you to use Fertilizer X. You plant 545 trees, and you give $n = 254$ of them Fertilizer X and $m = 291$ Fertilizer Y, and measure their height after three years.

   Now you have iid samples (assume trees grow independently) $x_1, x_2, ..., x_n$ which measure the height of the n trees given fertilizer X, and iid samples $y_1, y_2, ..., y_m$ which measure the height of the m trees given fertilizer Y.

   The data you are given has the following statistics:

Fertilizer	Number of samples	Sample Mean	Sample Variance
X	$n = 254$	$\bar{x} = 6.99$	$s_x^2 = 28.56^2$
Y	$m = 291$	$\bar{y} = 4.21$	$s_y^2 = 23.97^2$

   Perform a hypothesis test using the procedure in 8.4, and report the exact p-value for the observed difference in means. In other words: assuming that the heights of trees which had been given fertilizer X and fertilizer Y has the same mean μ_X, μ_Y, what is the probability that you could have sampled two groups of tress such that you could have observed that the difference of means between Fertilizer Y and Fertilizer X was as extreme, or more extreme, than the one observed (which is $\bar{x} - \bar{y} = 2.78$)?

   Solution:
   Our null and alternative are (since your friend claims that Fertilizer X is better than Y):

$$H_0 : \mu_X = \mu_Y \qquad H_A : \mu_X > \mu_Y$$

   Let's choose our significance level $\alpha = 0.05$. By the CLT, $\bar{X} \sim \mathcal{N}(\mu_x, s_x^2/n)$ and $\bar{Y} \sim \mathcal{N}(\mu_y, s_y^2/m)$. By closure properties of the normal and our null hypothesis (under this, $\mu_X = \mu_Y \rightarrow \mu_X - \mu_Y = 0$),

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu = 0, \sigma^2 = \frac{28.56^2}{254} + \frac{23.97^2}{291} = 5.18575\right)$$

   Then, we are asking

$$\begin{aligned} P(\bar{X} - \bar{Y} \geq \bar{x} - \bar{y}) &= P\left(\frac{(\bar{X} - \bar{Y}) - (\mu_{\bar{X}} - \mu_{\bar{Y}})}{\sqrt{5.18575}} \geq \frac{(\bar{x} - \bar{y}) - (\mu_{\bar{X}} - \mu_{\bar{Y}})}{\sqrt{5.18575}}\right) \\ &= P\left(\frac{(\bar{X} - \bar{Y}) - 0}{\sqrt{5.18575}} \geq \frac{(2.78) - 0}{\sqrt{5.18575}}\right) \\ &= P(Z \geq 1.22) = 0.1112 \end{aligned}$$

   Since our p-value of 0.1112 is $> \alpha = 0.05$, we fail to reject the null hypothesis. There is insufficient evidence to show that Fertilizer X is better than Fertilizer Y.