Unit 3

Team 8

April 7, 2018

Contents

Q53:

$$MSE(\hat{\theta}) = \text{Var}\left(\hat{\theta}\right) + Bias(\hat{\theta}, \theta)^2 = \text{Var}\left(\hat{\theta}\right) + (\mathbb{E}\left[\hat{\theta}\right] - \theta)^2$$

Case 1:

$$\hat{p_1} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$MSE(\hat{p_1}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) + (\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - p)^2 = \frac{p(1-p)}{n}$$

Case 2:

$$\hat{p_2} = \frac{1}{n+2}\sum_{i=1}^{n} X_i + 1$$

$$\begin{aligned} MSE(\hat{p_2}) &= \text{Var}\left(\hat{p_2}\right) + (\mathbb{E}\left[\hat{p_2}\right] - p)^2 = \\ &= \frac{1}{(n+2)^2}\sum_{i=1}^{n} \text{Var}\left(X_i\right) + (\frac{1}{n+2}\sum_{i=1}^{n} \mathbb{E}\left[X_i\right] + 1 - p)^2 = \\ &= \frac{np(1-p)}{(n+2)^2} + (\frac{np}{n+2} + 1 - p)^2 \end{aligned}$$

Case 3:

$$\hat{p_3} = X_1$$

$$MSE(\hat{p_3}) = \text{Var}\left(\hat{p_3}\right) + ((\mathbb{E}\left[\hat{p_3}\right] - p)^2) = p(1-p)$$

$$MSE(\hat{p_1}) < MSE(\hat{p_3}), \forall n > 1$$

$$\lim_{n\to\infty} MSE(\hat{p_2}) \to \frac{p(1-p)}{n} + 1 \implies MSE(\hat{p_1}) < MSE(\hat{p_2}), n \to \infty$$

So, $\hat{p_1}$ is uniformly better when $n \to \infty$.

Q54:

There are 2^n possible realizations of $(X_1, ..., X_n)$ Any statistic T takes string into a real number, t_j, where $j = 1, .., 2^n$.

Let's compute the expectation directly:

$$\mathbb{E}[T] = \sum_{j=1}^{2^n} t_j p^{n_j} (1-p)^{n-n_j}$$

here n_j is a number of successes.

For T to be unbiased we must have a condition:

$$\sum_{j=1}^{2^n} t_j p^{n_j} (1-p)^{n-n_j} = \frac{p}{1-p}, \forall p \in (0,1)$$

It is clear that this condition cannot be satisfied, hence it follows that no unbiased estimator of the odds ratio exists.

Q56:

We have shifted standard exponential distribution with density:

$$f(x|\theta) = e^{-(x-\theta)}, x \geq \theta$$

Part a:

$$lik(\theta) = \prod_{i=1}^{n} e^{-(X_i - \theta)}, X_1, ...X_n \geq \theta$$

Note also that $X_1, ...X_n \geq \theta$ is equivalent to $min(X_i) \geq \theta \implies$

$$l(\theta) = log(lik(\theta)) = -\sum_{i=1}^{n}(X_i) + n\theta, min(X_i) \geq \theta$$

It is obvious that the likelihood is maximized when θ attains its maximum value, which is $min(X_i)$

$$\hat{\theta}_{MLE} = min(X_i)$$

Part b: Is our estimate consistent?

Consistency means that $lim_{n\to\infty} MSE = 0$

$$MSE = E_\theta(\hat{\theta}_n - \theta)^2$$

We can notice that with $n \to \infty$ $\hat{\theta}_n$ is decreasing, and according to the condition that $\hat{\theta}_n = min(X_i)$ and at the same time we have condition that $X \geq \theta$, the low boundary will be for $\hat{\theta}_n = \theta$.

That is why,

$$lim_{n\to\infty} E_\theta(\hat{\theta}_n - \theta)^2 = 0$$

Consequently,our estimate $\hat{\theta}_n$ is consistent.

Part c: Distribution of $n(\hat{\theta}_n - \theta)$:

$F_{n(\hat{\theta}_n-\theta)} = P(n(\hat{\theta}_n - \theta) \leq x) = P(n\hat{\theta}_n \leq x + n\theta) = P(\hat{\theta}_n \leq \frac{x}{n} + \theta) = P(X_{min} \leq \frac{x}{n} + \theta)$.

So, we need to find cdf of X_{min}. For $X_1, X_2, ..., X_n$ iid continuous random variables with cdf of exponential (standard) shifted distribution, the cdf of minimum order statistic will be the following:

$F_{X_{min}}(x) = P(X_{min} \leq x) = 1 - P(X_{min} > x) = 1 - P(X1 > x, ..., Xn > x) = 1 - P(X1 > x)...P(Xn > x) = 1 - (1 - F(x))^n$

Going back to:

$F_{n(\hat{\theta}_n-\theta)} = P(X_{min} \leq \frac{x}{n} + \theta) = 1 - (1 - F(\frac{x}{n} + \theta))^n = 1 - (1 - 1 + e^{-(\frac{x}{n}+\theta-\theta)})^n = 1 - e^{-x}$

That is the cdf of standard exponential distribution.

Part d: A asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$

$F_{\sqrt{n}(\hat{\theta}_n-\theta)} = P(\sqrt{n}(\hat{\theta}_n - \theta) \leq x) = P(\hat{\theta}_n \leq \frac{x}{\sqrt{n}} + \theta) = 1 - e^{-x\sqrt{n}}$

Now if we approach n to infinity, we get that our asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ tends to 1.

Part e In the prove of the theorem "Assimptotic normality of MLE" we use the fact that $l'(\hat{\theta_n}) = 0$ but in shifted standard exponential distribution $l'(\hat{\theta_n}) = n \neq 0$. Normality of the MLE not hold in this case.

Q57:

Let there be N+1 urns each containing N balls such that the k-th urn contains k red and N-k blue balls (k = 0,1...N). At the first stage of the experiment we choose a random urn (with the probability of $\frac{1}{N+1}$ and then proceed to pick up balls from the chosen urn.

After the ball's color has been recorded, the ball is returned back to the urn. Assume the red ball showed up s times (event A) and - on the basis of that observation - predict the probability of the red ball showing up on the next trial (event B). We are thus looking into the conditional probability $P(B|A)$. Note that

$$P(B|A) = \frac{P(AB)}{P(A)}$$

For the urn with k red balls, the probability $P_k(s, n)$ of having s red balls in n trials is:

$$P_k(s, n) = \binom{n}{s} \frac{k^s(N-k)^{n-s}}{N^n}$$

In case, where s=n:

$$P_k(n, n) = (\frac{k}{N})^n$$

$P(AB) = P(B)$ making $P(B|A) = P(B)/P(A)$ so that what is needed is to evaluate the two probabilities P(A) and P(B) separately.

Define Q(k) as the probability of having a string of k successes in k trials. Then P(A)=Q(n) and P(B)=Q(n+1). But

$$Q(n) = \frac{1}{N+1} \sum_{k=0}^{N} \left(\frac{k}{N}\right)^n$$

which can be viewed as a Riemann sum approximation of the integral

$$\int_0^1 t^n dt = \frac{1}{n+1}$$

It follows that, for s=n

$$P(B|A) = \frac{Q(n+1)}{Q(n)} = \frac{n+1}{n+2}$$

Q58:

First we show that the gamma distribution is a conjugate prior for the exponential distribution. By a definition in the textbook we have that: If the prior distribution belongs to a family G and conditional on the parameters of G, the data have a distribution H then G is said to be conjugate to H if the posterior is in the family G.

So let $\theta \sim Gamma(\alpha, \beta)$ such that we have a prior:

$$f_\Theta(\theta) = \frac{\beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)}$$

And now we assume that the data given the parameter θ have an exponential distribution, we can write this as $X_i|\theta \sim exp(\theta)$

$$f_{X_i|\Theta}(x_i|\theta) = \theta e^{-\theta x_i}$$

Now the joint distribution of the n i.i.d samples is given by:

$$f(X|\theta) = \theta^n e^{-\theta \sum x_i}$$

We now use the following to get the posterior:

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

$$\Rightarrow f_{\Theta|X}(\theta|x) \propto f_{x|\Theta}(x|\theta) \times f_\Theta(\theta)$$

$$f_{\Theta|X}(\theta|x) \propto \theta^n e^{-\theta \sum x_i} \times \frac{\beta^\alpha \theta^{\alpha-1} e^{-\beta\theta}}{\Gamma(\alpha)}$$

Now since we are deriving based on a proportion we can eliminate any constants, since they too only add proportion (I don't know if that made sense, it did in my head), so we can write:

$$f_{\Theta|X}(\theta|x) \propto \theta^n e^{-\theta \sum x_i} \times \theta^{\alpha-1} e^{-\beta\theta}$$

$$= \theta^{n+\alpha-1} e^{-\theta(\sum x_i + \beta)}$$

$$\therefore \theta|X \sim Gamma(n+\alpha, \sum x_i + \beta)$$

and so a conjugate.

Now we consider the queue being modeled as an exponential. First we formulate the general form of the posterior then consider each of the two cases:

$$f_{\Lambda|X}(\lambda|x) \propto \lambda^{\sum x_i + \alpha - 1} e^{-(n+\beta)\lambda}$$

where we derive the parameters for the gamma as follows:

$$\beta = \mu_1/(\mu_2 - \mu_1^2) \ \text{ and } \ \alpha = \beta\mu_1$$

Case I: Mean of the gamma is .5 and $\sigma = 1$. Then we formulate $\mu_2 = \mu_1^2 + \sigma^2 = .5^2 + 1 = 1.25$, then we have that:

$$\beta = .5/(1.25 - .25) = .5 \ \text{ and } \ \alpha = .5 \times .5 = .25$$

It follows then that the posterior gamma has parameters:

$$\alpha' = \sum x_i + \alpha = 102 + .5 = 102.5$$

$$\beta' = 20 + .5 = 20.5$$

So posterior mean results in:

$$\mu_{post} = 102.5/20.5 = 5$$

Case II: Here we consider gamma mean to be 10 and standard deviation 20. Analogous to the above we can formulate the parameters of the priors as follows:

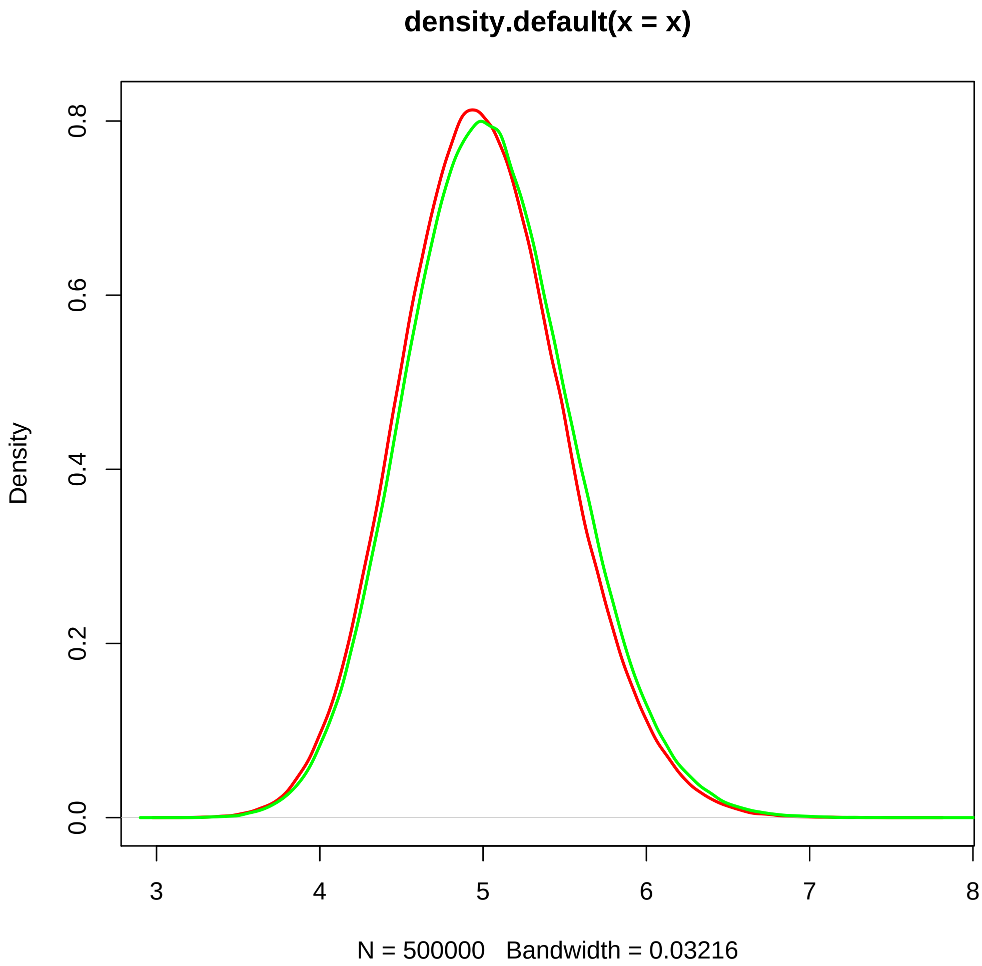
$$\beta = 10/(500 - 100) = .025 \ \text{ and } \ \alpha = .025 \times 10 = .25$$

$$\Rightarrow \alpha'' = 102 + .025 = 102.025 \ \text{ and } \ \beta'' = 20 + .25 = 20.25$$

And so posterior mean is:

$$\mu_{post} = 102.25/20.025 = 5.106$$

```r
x <- rgamma(500000,shape=102.5,rate = 20.5)
y <- rgamma(500000,shape=4081/40,rate = 81/4)
plot(density(x),col='red',lwd=2)
lines(density(y),col='green',lwd=2)
```



As we can see the gamma for the first case is a bit more condensed, moreover the posterior for the second case appears to be more centered about the true value.

Q59:

What is the posterior density of θ?

Note that our observed data (the number of successful shots), x given our parameter θ has a Binomial distribution:

$$x|\theta \sim Binomial(2, \theta)$$

since x is the result of two indpendent trials (we observed x = 2), each with probability of success θ. This implies that:

$$f(x|\theta) = \binom{2}{x}\theta^x(1-\theta^{2-x})$$

The Uniform[0,1] prior for θ implies $f(\theta) = 1$.
Normalizing constant is given by:

$$\int_0^1 p(x|\theta)p(\theta)d\theta = \int_0^1 \binom{2}{x}\theta^x(1-\theta^{2-x})d\theta = \binom{2}{x}B(x+1, 3-x)$$

where B(x,y) - beta function.
Thus the posterior is given by:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_0^1 p(x|\theta)p(\theta)d\theta} = \frac{\theta^x(1-\theta^{2-x})}{B(x+1, 3-x)}$$

Which corresponds to the Beta(x + 1; 3 - x) distribution

What would you estimate the probability that she makes a third shot to be?

Given that we have observed x = 2, the probability that she makes a third shot follows a Beta(3,1) distribution, and so we would estimate the probability that she makes a third shot to be the expected value of the Beta(3,1) distribution.

Since the $Beta(\alpha, \beta)$ distribution has expectation $\frac{\alpha}{\alpha+\beta}$, we would estimate the probability to be $\frac{3}{3+1} = \frac{3}{4}$.

Q60:

Part a:

In Bayesian probability theory, if the posterior distributions $p(\theta|x)$ are in the same family as the prior probability distribution $p(\theta)$, the prior and posterior are then called conjugate distributions.The usual conjugate prior is the beta distribution with parameters (α, β):

$prior(p|x) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$ We know that for optimist prior expectation is 1/2 with a prior variance of 1/36. Let's recall the mean and variance for the Beta distribution:

$$E(X) = \frac{\alpha}{\alpha+\beta} = \frac{1}{2}$$

$$2\alpha = \alpha + \beta$$

$$=> \beta = \alpha$$

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2} = \frac{1}{36}$$

$$\frac{\beta}{(2\beta+1)4\beta^2} = \frac{1}{36}$$

$$8\beta + 4 = 36$$

$$=> \beta = 4 = \alpha$$

So, for optimist prior has Beta distribution:

$$prior(p|x) = \frac{p^3(1-p)^3}{B(4,4)}$$

Important to notice that if we put in the density of Beta distribution parameters $\alpha = 1, \beta = 1$, we will get the following:

$$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} = \frac{1}{B(\alpha,\beta)} = \frac{1}{B(1,1)} = 1.$$

And this is the density of $U(0,1)$. Recall: $f_U = 1/b - a$, when $a \leq x \leq b$. Plug in $a = 0$, $b = 1$, we get our result. From here we can conclude that Uniform distribution belongs to family of Beta distributions.

Part B:
Jeffrey's prior is proportional to the square root of the determinant of the Fisher information:

$$p\left(p\right) \propto \sqrt{\det \mathcal{I}\left(p\right)}$$

.

We know that gambling is the a Bernoulli experiment (with probability p you win, with $(1-p)$ you lose). But since we repeat this experiment n times, we get the Binomial distribution. So we need to find I_p:

$$f(x) = p^k(1-p)^{n-k}$$

$$l = k\log p + (n-k)\log(1-p)$$
$$l' = \frac{k}{p} - \frac{n-k}{1-p}$$
$$l'' = -\frac{k}{p^2} + \frac{n-k}{(1-p)^2}$$
$$I_p = -E(\frac{k}{p^2} + \frac{n-k}{(1-p)^2}) = -(\sum_{k=0}^{n}(-\frac{k}{p^2} + \frac{n}{(1-p^2)} - \frac{k}{(1-p^2)})\binom{n}{k}p^k(1-p)^{n-k} =$$
$$= -\frac{n}{p} - \frac{np}{(1-p^2)} + \frac{n}{1-p^2} = \frac{n}{p(1-p)}$$

So our prior is the following:

$$\sqrt{\frac{n}{p(1-p)}} = n^{1/2}p^{-1/2}(1-p)^{-1/2}$$

And here we can notice that this belongs to the family of Beta distribution with parameters (1/2, 1/2).

Part c:

We know that if prior has distribution from Beta family, and as we have already showed gambling has to the Binomial distribution, so we need to end up with posterior from Beta distribution as well but we need to add number of success (12) to parameter α and number of fails (13) to β. So, we obtain our posterior:

Optimist : Beta(16,17)
Realist: Beta(13,14)
Pessimist: Beta(25/2, 27/2)

Part d:

Posterior means (using formula for mean of Beta distribution but with new parameters:

Optimist : $\frac{16}{16+17} = 48.5\%$
Realist: $\frac{13}{13+14} = 48,2\%$
Pessimist: $\frac{25/2}{25/2+27/2} = 48,1\%$

Part e:

From MLE estimate we know that random variable $\sqrt{nI_p}(\hat{p}_n - p) \sim N(0,1)$. We already calculated above:

$$I_p = \frac{n}{p(1-p)} = \frac{25}{12/25(1-13/25)} = 100.16$$

$$\hat{p} = (12\log(p) + 13\log(1-p))' = 0$$

$$=> 12/p - 13/(1-p) = 0 => \hat{p} = 12/25$$

So, our confidence interval will be:

$$(\hat{p} - \frac{z}{\sqrt{nI}}, \hat{p} + \frac{z}{\sqrt{nI}}).$$

We know that z we take from Laplace table and for confidence level 95% it will be 1.96. Hence:

$$(0.48 - \frac{1.96}{\sqrt{25 * 100.16}}, 0.48 + \frac{1.96}{\sqrt{25 * 100.16}})$$

$$(0.44, 0.52)$$