Stat 513
Homework key 7

March 11, 2016

REGULAR PROBLEMS

1. Suppose that $Y_1, Y_2, \ldots, Y_n$ are i.i.d. with density
$$f_Y = e^{-(y-\mu)} 1_{(\mu,\infty)}$$
for $\mu \in (-\infty, \infty)$. Use Basu's theorem to show that $Y_{(1)}$ is independent of the sample variance s^2.
Ans: Notice that $X_i = Y_i - \mu \sim \mathcal{E}(1)$. Clearly S_X^2 does not depend on μ. Since $S_X^2 = S_Y^2 = s^2$, hence s^2 is ancillary. Therefore, it is enough to show that $Y_{(1)}$ is complete and minimal sufficient.
$$f(y_1, \ldots, y_n|\mu) = e^{-\sum_{i=1}^n y_i + n\mu} 1_{[y_{(1)} > \mu]}$$
Therefore the minimal sufficient statistics is $Y_{(1)}$. Recall
$$Y_{(1)} = X_{(1)} + \mu$$
and $X_{(1)}$ is the minimum of n $\mathcal{E}(1)$ random variables. So
$$X_{(1)} \sim \mathcal{E}(n),$$
and clearly
$$Y_{(1)} \sim e^{-n(y-\mu)} 1_{[y_{(1)} > \mu]}.$$
Let h be a real function. Let
$$\begin{aligned} &Eh(Y_{(1)}) = 0 \\ &\Rightarrow \int_\mu^\infty h(y) e^{-n(y-\mu)} = 0 \\ &\Rightarrow \int_\mu^\infty h(y) e^{-ny} = 0 \end{aligned}$$
for all $\mu > 0$. Taking derivative w.r.t. μ we get that
$$0 - h(\mu) e^{-n\mu} = 0, \forall \mu > 0.$$
Therefore
$$h(\mu) = 0 \forall \mu > 0$$
leading to $h = 0$. Hence, $Y_{(1)}$ is complete.

2. Suppose that for data $Y_1, \ldots, Y_n$ conditional on the parameter p, have distribution as i.i.d. Bernoulli random variables with $Y_i|p \sim \mathcal{B}(1, p)$, $p \in (0, 1)$. Further suppose prior distribution $p \sim Beta(\alpha, \beta)$.

(a) Show that this prior distribution is the conjugate prior distribution.

(b) Find the posterior distribution of p conditional on $\overrightarrow{Y} = (Y_1, \ldots, Y_n)$.

(c) Find the Bayes estimator of p based on squared error loss. Can this estimator be expressed as a weighted average of the data and the prior distribution? If so, do so.

(d) Can you find a closed form solution for the Bayes estimator of p based on absolute error loss. If so, do so. If not, explain why not.

Ans:

(a)

$$
\begin{aligned}
f(\overrightarrow{y}) &= p^{\sum_{i=1}^n y_i}(1-P)^{n-\sum_{i=1}^n y_i} \\
f(p, \overrightarrow{y}) &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} p^{\alpha-1}(1-p)^{\beta-1} \\
f(p|\overrightarrow{y}) &\sim p^{\sum_{i=1}^n y_i+\alpha-1}(1-p)^{n-\sum_{i=1}^n y_i+\beta-1} \\
\Rightarrow p|\overrightarrow{Y} &\sim Beta(\sum_{i=1}^n y_i+\alpha, n-\sum_{i=1}^n y_i+\beta)
\end{aligned}
$$

Hence, the prior is a conjugate prior.

(b)

$$
p|\overrightarrow{Y} \sim Beta(\sum_{i=1}^n y_i+\alpha, n-\sum_{i=1}^n y_i+\beta)
$$

(c) The Bayes estimator is

$$
\hat{p} = E[p|\overrightarrow{Y}] = \frac{\sum_{i=1}^n Y_i+\alpha}{n+\alpha+\beta}
$$

It is possible to express $\hat{p}$as a weighted average of the data and the prior distribution.

$$
\begin{aligned}
\hat{p} &= \frac{\sum_{i=1}^n Y_i}{n}\frac{n}{\alpha+\beta+n} + \frac{\alpha}{\alpha+\beta}\frac{\alpha+\beta}{\alpha+\beta+n} \\
&= \bar{Y}\frac{n}{\alpha+\beta+n} + Ep\frac{\alpha+\beta}{\alpha+\beta+n}
\end{aligned}
$$

(d) The Bayes estimator of p based on absolute error loss is going to be the median of $Beta(\sum_{i=1}^n y_i+\alpha, n-\sum_{i=1}^n y_i+\beta)$. $Beta$ distribution does not allow for a closed form solution for the median.

3. Suppose that for data $Y_1, Y_2, \ldots, Y_n$, conditional on the hazard parameter λ, have distribution as i.i.d. exponential random variables with $Y_i|p \sim \mathcal{E}(\lambda)$, $\lambda \in (0, \infty)$. Further suppose prior distribution $\lambda \sim Gamma(\alpha, \beta)$.

(a) Show that this prior distribution is the conjugate prior distribution.

(b) Find the posterior distribution of λ conditional on $\vec{Y} = (Y_1, \ldots, Y_n)^T$.

(c) Find the Bayes estimator of λ based on squared error loss. Can this estimator be expressed as a weighted average of the data and the prior distribution? If so, do so.

(d) Can you find a closed form solution for the Bayes estimator of λ based on absolute error loss? If so, do so. If not, explain why not.

Ans:

(a) We have that for a single observation, $f(y_i|\lambda) = \lambda e^{-\lambda y_i}$. Hence, we have the following likelihood and prior:

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} f(y_i|\lambda) \\ &= \lambda^n \exp(-\lambda \sum_{i=1}^{n} y_i) \\ \pi(\lambda) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}. \end{aligned}$$

We can find the posterior distribution of $\lambda|\vec{Y}$ as follows,

$$\begin{aligned} \pi(\lambda|\vec{Y}) &= \frac{L(\lambda)\pi(\lambda)}{f(\vec{y})} \\ &\propto L(\lambda)\pi(\lambda) \\ &= \lambda^n \exp(-\lambda \sum_{i=1}^{n} y_i) \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \\ &\propto \lambda^{\alpha+n-1} \exp\{-(\beta + \sum_{i=1}^{n} y_i)\}. \end{aligned}$$

We now recognize this as the kernel of a *Gamma* pdf. Hence we are using the conjugate prior for this family.

(b) In particular, we note from (a) that the posterior distribution is

$$(\lambda|\vec{Y}) \sim \Gamma(\alpha + n, \beta + \sum_{i=1}^{n} Y_i).$$

(c) The Bayes estimator of λ based on squared error loss is the posterior mean. Here, we find that

$$\begin{aligned} E[\lambda|\vec{Y}] &= \frac{\alpha + n}{\beta + \sum_{i=1}^{n} Y_i} \\ &= \frac{\beta}{\beta + \sum_{i=1}^{n} Y_i} \frac{\alpha}{\beta} + \frac{\sum_{i=1}^{n} Y_i}{\beta + \sum_{i=1}^{n} Y_i} \frac{n}{\sum_{i=1}^{n} Y_i} \\ &= w_0 E[\lambda] + w_1 \hat{\lambda}, \end{aligned}$$

so that we express the posterior mean as a weighted average of the prior mean and MLE $\hat{\lambda}$ of λ.

(d) The Bayes estimator of λ based on absolute error loss is the posterior median, $\tilde{\lambda}$, which satisfies

$$\frac{1}{2} = \int_0^{\tilde{\lambda}} \frac{(\beta + \sum Y_i)^{\alpha+n}}{\Gamma(\alpha+n)} \lambda^{\alpha+n-1} e^{-(\beta+\sum Y_i)\lambda} d\lambda.$$

Unfortunately, we do not have general closed form solutions for the median of a gamma distribution. Calculation of the posterior median should thus be done numerically.

4. Suppose $Y_1, \ldots, Y_n$ are i.i.d. gamma random variables with $Y_i \sim \Gamma(\gamma, \lambda)$ with $\gamma > 0$ known and $\lambda > 0$ the rate parameter (so $E[Y_i] = \gamma/\lambda$.)

(a) Find the most powerful level α test of the simple hypotheses $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda = \lambda_1 < \lambda_0$. Show that the test is a function of the sufficient statistic, and explicitly provide a definition of the critical value. Is the test unbiased? Is the test consistent?

(b) Extend the test of part a to the test of the composite hypotheses $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda = \lambda_1 < \lambda_0$. Show that the power function is decreasing in λ. Is the test unbiased? Is the test consistent?

(c) Define a two-sided $100(1-\alpha)\%$ confidence interval for λ.

Ans:

(a) First construct the MP test for $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda = \lambda_1$ where $\lambda_1 < \lambda_0$. By Neyman Pearson lemma, the MP test $\phi(\overrightarrow{Y}) = 1$ if

$$T = \frac{f_{\lambda_1}(\overrightarrow{Y})}{f_{\lambda_0}(\overrightarrow{Y})} = \frac{\exp(-\lambda_1 \sum_{i=1}^n Y_i)\lambda_1^{n\gamma}}{\exp(-\lambda_0 \sum_{i=1}^n Y_i)\lambda_0^{n\gamma}} > K_0$$

Notice that since $\lambda_1 < \lambda_0$, T is increasing in $\sum_{i=1}^n Y_i$. Therefore the above is equivalent to

$$\sum_{i=1}^n Y_i > K$$

$\phi = 0$ o.w. Now $P_{\lambda_0}(\sum_{i=1}^n Y_i > K) = \alpha$. $\sum_{i=1}^n Y_i \sim \Gamma(n\gamma, \lambda_0)$ under the null. Therefore $K = \Gamma_{1-\alpha, n\gamma, \lambda_0}$, the $1-\alpha$-th quantile of $\Gamma(n\gamma, \lambda_0)$. We observe that the test does not depend on λ_1. It only depends on the fact that $\lambda_1 < \lambda_0$. Therefore ϕ is MP for each $\lambda_1 < \lambda_0$. Hence, ϕ is also UMP for the original test $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda = \lambda_1 < \lambda_0$.

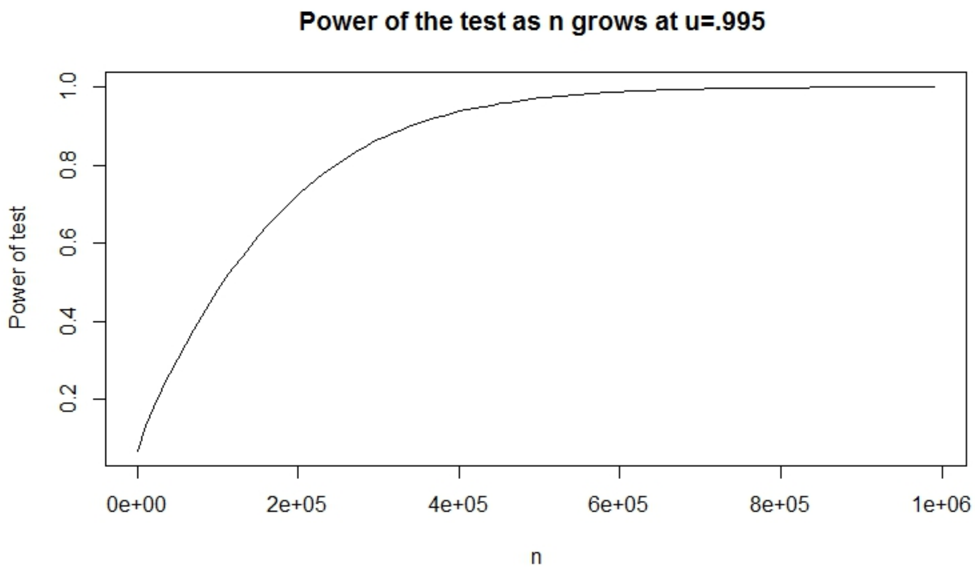
We will show that the power function of this test is a decreasing function λ. Before that, we notice that $X_i = \lambda Y_i \sim Gamma(\gamma, 1)$ if $Y_i \sim Gamma(\gamma, 1)$. By the same argument, we also get that

$$\lambda_0 K = \lambda_0 \Gamma_{1-\alpha, n\gamma, \lambda_0} = \Gamma_{1-\alpha, n\gamma, 1} = C_n.$$

Therefore

$$\beta(\lambda) = P_\lambda(\sum_{i=1}^n Y_i > K) = P_\lambda(\sum_{i=1}^n X_i > \frac{\lambda}{\lambda_0} C_n). \quad (1)$$

X_i, C_n does not depend on λ. As λ increases, clearly $\beta(\lambda)$ decreases. Hence $\inf_{\lambda \leq \lambda_0} \beta(\lambda) \geq \beta(\lambda_0) = \alpha$. Therefore, the test is unbiased. To check if the test is consistent, we took $\gamma = 1$. For large n, the value of γ will not matter so WLOG we can take $\gamma = 1$. Then we simulated random variable S_n from $Gamma(n, 1)$. Now for different values of λ, λ_0, $\dfrac{\lambda}{\lambda_0}$ is going to be a fraction. If we can show that for any fraction u, $P(S_n > uC_n) \to 1$ we will be able to establish that the test is consistent. Observe if for some u, $P(S_n > uC_n) \to 1$ then the same holds for all $u' < u$. Therefore we can take f to be a fraction very close to 1. we took $u = .995$ and calculated $P(S_n > .995C_n)$. We saw that the probability approaches 1. Hence, we conclude that the test is consistent.



(b) Our claim is that the test ϕ in part (a) is the UMP level α test for part (b). The test is level-α by (1). To see why, observe that since $P_\lambda(\sum_{i=1}^n Y_i > K)$ is decreasing in λ, $\sup_{\lambda \geq \lambda_0} P_\lambda(\sum_{i=1}^n Y_i > K) \leq P_{\lambda_0}(\sum_{i=1}^n Y_i > K) = \alpha$. Therefore the size of the test is α. Hence, the test is level α. Now, if a test is level α for part (b), it is also level α for part (a). Since the test ϕ is UMP among all level α tests for part (a), it is also UMP among all level α tests for part (b), precisely a subclass of the former. we have already shown that the power function is decreasing and this test is unbiased and consistent.

(c)

$$P(\Gamma_{\alpha/2,n\gamma,1} \le \sum_{i=1}^{n} X_i \le \Gamma_{1-\alpha/2,n\gamma,1}) = 1-\alpha$$

$$\Rightarrow P(\Gamma_{\alpha/2,n\gamma,1} \le \lambda \sum_{i=1}^{n} Y_i \le \Gamma_{1-\alpha/2,n\gamma,1}) = 1-\alpha$$

$$\Rightarrow P(\frac{\Gamma_{\alpha/2,n\gamma,1}}{\sum_{i=1}^{n} Y_i} \le \lambda \le \frac{\Gamma_{1-\alpha/2,n\gamma,1}}{\sum_{i=1}^{n} Y_i}) = 1-\alpha.$$

therefore $(100-\alpha)\%$ C.I. is $(\frac{\Gamma_{\alpha/2,n\gamma,1}}{\sum_{i=1}^{n} Y_i}, \frac{\Gamma_{1-\alpha/2,n\gamma,1}}{\sum_{i=1}^{n} Y_i})$.

5. Suppose $Y_1, Y_2, \ldots, Y_n$ are independent random variables with $Y_i \sim (\mu_i, \sigma^2)$ and the distribution is otherwise unspecified. Further suppose that for known $x_1, \ldots, x_n$, $\mu_i = \beta_0 + \beta_1 x_i$. We are interested in testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$.

   (a) First presuming that σ^2 is known, derive an approximate level α test of H_0 versus H_1, providing an explicit expression for the critical value. Be sure to explain the conditions under which your test will be a good approximation to a level α test.

   (b) Show that your test in part a corresponds to a likelihood ratio test (LRT) when the Y_i's satisfy $Y_i|x_i \sim \mathcal{N}(\mu_i, \sigma^2)$.

   (c) Now modify your test in part a to the case where σ^2 is unknown.

   (d) How would your test in part c differ from a test derived when the data were known to be normally distributed? Are the tests asymptotically equivalent?

Ans:

   (a) Without loss of generality, suppose that $x_1, \ldots, x_n$ are centered covariates such that $\bar{x} = 0$. Recall that the OLS estimate $\hat{\beta}_1$ is approximately normal for large n with an approximate distribution given by

$$\hat{\beta}_1 = \frac{\sum x_i Y_i}{\sum x_i^2} \dot{\sim} \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum x_i^2}\right).$$

This approximation is valid as long as we can assume that

$$\lim_{n\to\infty} \frac{\sup_i x_i^2}{\sum x_i^2} = 0.$$

Assuming the normal approximation is valid, we have further that under H_0,

$$\frac{\sum x_i^2}{\sigma^2} \hat{\beta}_1^2 \dot{\sim} \chi_1^2.$$

This leads to an approximate level α test given by

$$\phi(\vec{Y}) = \mathbf{1}_{[\chi^2_{1,1-\alpha},\infty)}\left(\frac{\sum x_i^2}{\sigma^2}\hat{\beta}_1^2\right),$$

where $Pr(\chi_1^2 > \chi^2_{1,1-\alpha}) = \alpha$.

(b) Recall that the LRT is given by $\phi(\vec{Y}) = \mathbf{1}[\Lambda(\vec{Y}) > k]$, with

$$\Lambda(\vec{Y}) = \frac{\sup_{\Theta_1} L(\vec{\theta}|\vec{Y})}{\sup_{\Theta_0} L(\vec{\theta}|\vec{Y})}.$$

In the present case, the numerator obtains its supremum value at the MLE (and OLSE/BLUE/UMVUE) $\hat{\vec{\beta}}$. The denominator obtains its supremum value at the MLE subject to the constraint $\beta_1 = 0$, which is simply the vector $(\hat{\beta}_0, 0)$. Now,

$$\begin{aligned}
\Lambda(\vec{Y}) &= \frac{\sup_{\Theta_1} L(\vec{\theta}|\vec{Y})}{\sup_{\Theta_0} L(\vec{\theta}|\vec{Y})} \\
&= \frac{\prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\{-(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2/2\sigma^2\}}{\prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\{-(Y_i - \hat{\beta}_0)^2/2\sigma^2\}} \\
&= \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^n [(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 - (Y_i - \hat{\beta}_0)^2]\right) \\
&= \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^n [\hat{\beta}_1^2 x_i^2 - 2\hat{\beta}_1 x_i (Y_i - \hat{\beta}_0)]\right) \\
&= \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^n [\hat{\beta}_1^2 x_i^2 - 2\hat{\beta}_1 x_i (Y_i - \bar{Y})]\right) \\
&= \exp\left(\frac{-1}{2\sigma^2}\hat{\beta}_1^2 [\sum_{i=1}^n x_i^2 - 2\sum_{i=1}^n x_i^2]\right) \\
&= \exp\left(\frac{1}{2}\frac{\sum_{i=1}^n x_i^2}{\sigma^2}\hat{\beta}_1^2\right).
\end{aligned}$$

We note that under the null hypothesis, we have the exact distribution

$$2\log\Lambda = \frac{\sum_{i=1}^n x_i^2}{\sigma^2}\hat{\beta}_1^2 \sim \chi_1^2.$$

So, our test in a is equivalent to the level α test function

$$\phi(\vec{Y}) = \mathbf{1}_{[\chi^2_{1,1-\alpha},\infty)}(2\log\Lambda)$$

when we know that $Y_i|x_i \sim \mathcal{N}(\mu_i, \sigma^2)$.

(c) We define $S^2 = \frac{1}{n-2}\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$ to be an unbiased and consistent estimator of σ^2. Appealing to Slutsky's theorem, our χ_1^2 approximation used in a remains valid replacing σ^2 with S^2 so that our test function becomes

$$\phi(\vec{Y}) = \mathbf{1}_{[\chi^2_{1,1-\alpha},\infty)}\left(\frac{\sum x_i^2}{S^2}\hat{\beta}_1^2\right).$$

(d) When $Y_i|x_i \sim \mathcal{N}(\mu_i, \sigma^2)$, then

$$S^2 \sim \frac{\sigma^2}{n-2}\chi^2_{n-2}.$$

Hence, we find that the exact distribution of the statistic is

$$\frac{\sum x_i^2}{S^2}\hat{\beta}_1^2 = \frac{\frac{\sum_{i=1}^n x_i^2}{\sigma^2}\hat{\beta}_1^2}{S^2/\sigma^2} \sim F_{1,n-2} \to_d \chi_1^2,$$

which follows from problem 1 on HW 2. Hence, the test will be asymptotically equivalent to the test in c.

6. Suppose $Y_1, \dots, Y_n$ are i.i.d. Bernoulli random variables with $Y_i \sim \mathcal{B}(1, p)$.

   (a) Derive an expression for the upper $(100 - \alpha)\%$ confidence bound for p using the exact distribution of $\overrightarrow{Y}$.

   (b) Suppose we observe $Y_i = 0$ for all i. Show that for large n, the bound found in part a is approximately 3/n when $\alpha = .05$ and approximately 3.69/n when $\alpha = 0.025$

Ans:

(a) Denote $Y = \sum_{i=1}^n Y_i$, and $y = \sum_{i=1}^n y_i$. Let the C.I. be $(0, p_{UB})$. Then $p_{UB} = \sup\{p | P_p(Y \leq y) \geq \alpha\}$. therefore, p_{UB} must satisfy

$$P_p(Y \leq y) = \alpha.$$

It is possible to have a tractable expression for the above probability. to see how, we need to establish a relationship between Binomail and Beta distribution. We will now define Y_i in a different way. Suppose $X_i \sim U(0, 1)$. let $Y_i = 1$ if $X_i > 1 - p$. therefore $Y_i \sim \mathcal{B}(1, p)$. So the this definition of Y_i is valid. Now the k-th order statistics of $X_i, \dots, X_n$ has a Beta distribution. Actually,

$$X_{(k)} \sim Beta(k, n + 1 - k) \tag{2}$$

Notice

$$\begin{aligned}
&P_p(Y \leq y) \\
&= 1 - P_p(Y \geq y + 1) \\
&= 1 - P(\text{There at least y } + 1\ X_i\text{-s that are greater than } 1 - p) \\
&= 1 - P(\text{Largest } y + 1 \text{ many } \ X_i\text{-s are greater than } 1 - p) \\
&= 1 - P_p(X_{(n-y)} \geq 1 - p)
\end{aligned}$$

Now,

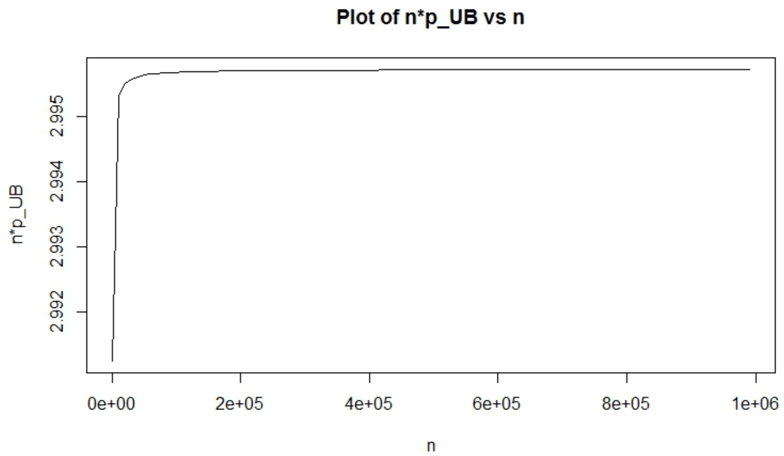
$$\begin{aligned}
&P_{p_{UB}}(Y \leq y) = \alpha \\
&\Rightarrow \alpha = P(X_{(n-y)} < 1 - p_{UB}) \\
&1 - p_{UB} = Betainv(\alpha, n - y, 1 + y)
\end{aligned}$$

since $X_{(n-y)} \sim Beta(n - y, 1 + y)$ by (2). Therefore a $(100 - \alpha)\%$ C.I. will be $(0, 1 - Betainv(\alpha, n - y, 1 + y))$.

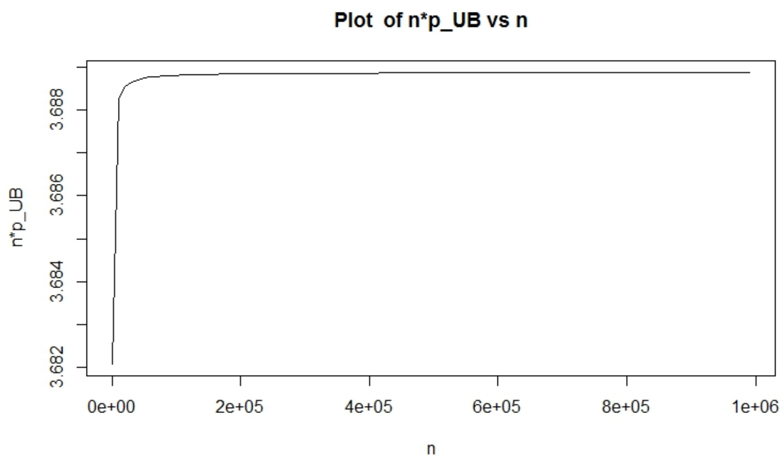
(b) We get $y = 0$. Let

$$\begin{aligned} x &= Betainv(\alpha, n, 1) \\ \Rightarrow \alpha &= \frac{n!}{n-1!} \int_0^x y^{n-1} dy \\ &= x^n \\ \therefore x &= (\alpha)^{1/n} \\ \therefore p_{UB} &= 1 - (\alpha)^{1/n}. \end{aligned}$$

We plot $n(1 - .05^{1/n})$ vs n and we see that the value approaches 3.69 as n becomes large.



We plot $n(1 - .025^{1/n})$ vs n and we see that the value approaches 3 as n becomes large.



MORE INVOLVED PROBLEMS