Brief solution to the regular final exam of ATS, PhD program in MAEM, Jan/24/2014.

1. (a) Let $\varepsilon > 0$. Then $P(|X_{(1)} - a| > \varepsilon) = P(X_{(1)} > a + \varepsilon) + P(X_{(1)} < a - \varepsilon)$. The last probability is zero. The first is $[a/(a+\varepsilon)]^{bn}$ which goes to zero, and this proves the result.

   (b) Let $z > 0$. Then $P[(X_{(n)} - a_n)/b_n \le z] = P(X_{(n)} \le b_n z - a_n)$ assuming $b_n > 0$. This can be written as

$$\left[1 - \left(\frac{a}{b_n z + a_n}\right)^b\right]^n$$

   and we want to find a_n and b_n such that this converges to $e^{-z^{-b}}$. This is the limit, for instance, of the sequence $[1 - z^{-b}/n]$ which can be obtained by setting $a_n = 0$ and $b_n = a\ n^{1/b}$. We can now go back and make sure that all calculations are fine as long as $z > 0$.

   (c) The central limit theorem states that $\sqrt{n}(\bar{X} - 3a/2)$ converges in distribution to $\mathrm{N}(0, 3a^2/4)$. This is equivalent to

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} \mathrm{N}(0, h(\mu))$$

   where $h(\mu) = \mu^2/3$. When applying the delta method, we want to find $g(\cdot)$ satisfying $[g'(\mu)]^2 = 1/h(\mu)$ that is, $g'(\mu) = \sqrt{3}/\mu$. Hence, $g(x) = \sqrt{3}\ \ln(x)$, $x > 0$.

2. (a) Consider two samples of size n, (x_1, y_1) and (x_2, y_2). It is easy to see that

$$\frac{f(x_1, y_1 \mid \theta)}{f(x_2, y_2 \mid \theta)} = \exp\left[\sum_i (x_{2i} - x_{1i})\ \theta\right]\ \exp\left[\sum_j (y_{2j} - y_{1j})/\theta\right]$$

   and this does not depend on θ iff $\sum_i x_{2i} = \sum x_{1i}$ and $\sum_j y_{2j} = \sum_j y_{1j}$. Hence, T is minimal suffient. When a most efficient estimator exists, it is sufficient. Since the minimal sufficient statistic is 2-dimensional, there cannot exist 1-dimensional statistics which are sufficient. Hence, there are no most efficient estimators of θ.

   (b) It is clear that $T_1 \sim \mathrm{G}(n, \theta)$ and $T_2 \sim \mathrm{G}(n, 1/\theta)$. Hence, $E[T_1\ T_2] = E[2\theta T_1\ 2T_2/\theta]/4 = (2n)^2/4 = n^2$ by independence. Hence, $T_1\ T_2 - n^2$ is a function of T which has zero expected value but is not identically zero, and this shows that T is not complete. In this case, there are no complete and sufficient statistics as either all minimal sufficient are complete, or there are no minimal sufficient statistics which are complete.

   (c) The likelihood is $L(\theta \mid x, y) \propto \exp(-\theta \sum x_i - \sum y_j/\theta)$ and it's straightforward to see that the value of θ that maximizes $\ln L$ is precisely t_2/t_1. This shows the result.

   (d) $E[T_2/T_1] = E[T_2]E[1/T_1]$ by independence. We already know that $E[T_2] = n\theta$. To obtain $E[1/T_1]$ one computes the integral

$$\int_0^{+\infty} 1/t\ f(t)dt$$

   where $f(t)$ is the density of a $\mathrm{G}(n, \theta)$ distribution by recognizing the kernel of a $\mathrm{G}(n-1, \theta)$. The end result is $\theta/(n-1)$. As such, $E[U] = \theta^2 n/(n-1)$, and hence U is a biased estimator of θ^2. $V = (n-1)U/n$ is thereofore and unbiased estimator of θ^2. (Alternatively, we can construct an F-Snedecor using T_1 and T_2.)

   (e) Clearly, T_2/n converges in probability to θ by the weak law of large numbers, and hence in distribution. By the same reasoning, T_1/n converges in probability to $1/\theta$ and, by the continuous mapping theorem, n/T_1 converges in probability to θ. Apply Slutsky to conclude that $U = (T_2/n) \times (n/T_1)$ converges in distribution to θ^2, and hence in probability, because convergence in probability to a constant is equivalent to convergence in distribution to a constant.

3. (a) For one observation of the model, we have $L(\theta \mid x) \propto \theta(1-\theta)^{x-1}$, $\ln L(\theta \mid x) = c + \ln\theta + (x-1)\ln(1-\theta)$, $d\ln L(\theta \mid x)/d\theta = 1/\theta - (x-1)/(1-\theta)$, $d^2 \ln L(\theta \mid x)/d\theta^2 = -1/\theta^2 - (x-1)/(1-\theta)^2$, and hence the Fisher information about θ in one observation is $I_X(\theta) = 1/\theta^2 + (1/\theta - 1)/(1-\theta)^2 = 1/[\theta^2(1-\theta)]$. The Jeffreys prior is thus $\pi(\theta) \propto \theta^{-1}\ (1-\theta)^{-1/2}$.

   For n observations, the likelihood is $L(\theta \mid x_1, \ldots, x_n) \propto \theta^n\ (1-\theta)^{t-n}$, where $t = \sum x_i$. Following Bayes theorem,

$$\begin{aligned}\pi(\theta \mid x_1, \ldots, x_n) &\propto L(\theta \mid x_1, \ldots, x_n)\\ &\propto \theta^n\ (1-\theta)^{t-n}\ \theta^{-1}\ (1-\theta)^{-1/2}\\ &\propto \theta^{n-1}\ (1-\theta)^{t-n-1/2}\\ &\propto \mathrm{Be}(\theta \mid n, t-n+1/2)\ .\end{aligned}$$

(b) The posterior mean is $n/(t+1/2)$, whereas the method of moments estimate is n/t. If θ is large and n is small, it's not unlikely that $t=n$, which will result in an unreasonable estimate of 1 for the probability of success using the method of moments estimate. This will not happen using the posterior mean.

(c) $E[X_{n+1} \mid x_1, \ldots, x_n] = E[E[X_{n+1} \mid \theta] x_1, \ldots, x_n]$. Since $E[X_{n+1} \mid \theta] = 1/\theta$, we have to compute

$$\begin{aligned}
E[1/\theta \mid x_1, \ldots, x_n] &= \int_0^1 \theta^{-1} \mathrm{Be}(\theta \mid n, t-n+1/2)\, d\theta \\
&= \frac{B(n-1, t-n+1/2)}{B(n, t-n+1/2)} \\
&= \frac{\Gamma(t+1/2)}{\Gamma(n)\,\Gamma(t-n+1/2)} \; \frac{\Gamma(n-1)\,\Gamma(t-n+1/2)}{\Gamma(t-1+1/2)} \\
&= \frac{t-1/2}{n-1}
\end{aligned}$$

where we have used the fact $\Gamma(t) = (t-1)\Gamma(t-1)$.