Solutions to HW2, Stat 401 Spring 2011

(1). #20, p. 251. (a). In this problem, we maximize over p the binomial probability mass function evaluated at X, by calculating

$$\frac{d}{dp} \log \Big(\binom{n}{X} p^X (1-p)^{n-X} \Big) = \frac{X}{p} - \frac{n-X}{1-p} = \frac{X-np}{p(1-p)}$$

which is decreasing in p and is 0 only when $p = \hat{p} = X/n$. So here $\hat{p} = 3/20$. (b) Since the binomial expectation is $E(X) = np$, we have $E(\hat{p}) = E(X/n) = p$, and the MLE is unbiased in this case. (c) The MLE of a function of the parameter is always obtained as the same function of the MLE, so in this case the MLE of $(1-p)^5$ is $(1-.15)^5 = .4437$.

(2). #28, p. 251. (a). Now we are to maximize the log-likelihood

$$\log \Big(\prod_{i=1}^{n} (X_i\, \vartheta^{-1}\, e^{-X_i^2/(2\vartheta)} \Big) = \sum_{i=1}^{n} \log(X_i) - n \log \vartheta - \frac{1}{2\vartheta} \sum_{i=1}^{n} X_i^2$$

The second derivative in ϑ is always negative, and the first derivative is $-n/\vartheta \sum_{i=1}^n X_i^2/(2\vartheta^2)$, which is 0 only when $\vartheta = \hat{\vartheta} = (2n)^{-1} \sum_{i=1}^n X_i^2$. For the data in Exercise 15, $\sum_{i=1}^n X_i^2 = 1490.11$, with $n = 10$, so $\hat{\vartheta} = 1490.11/20 = 74.51$. Since $E(X_1^2) = 2\vartheta$, the unbiased estimator in Exercise 15 was exactly the same as the estimator found here. [That is all this part of the exercise asked you to do. But in Ex. 15, since $E(X_1) = E(\bar{X}) = \sqrt{\pi\vartheta/2}$, the method-of-moments estimator $2(\bar{X})^2/\pi$ which on this set of data was equal to 81.29.]

(b). For this type of Rayleigh random variable, we find the median $m = m(\vartheta)$ as a function of ϑ by solving

$$\frac{1}{2} = F(m, \vartheta) = \int_0^m \frac{x}{\vartheta}\, e^{-x^2/(2\vartheta}\, dx = \int_0^{m/\sqrt{\vartheta}} y\, e^{-y^2/2}\, dy = 1 - e^{-m^2/(2\vartheta)}$$

This implies $m(\vartheta) = (2\vartheta \ln 2)^{1/2}$, and as before we find the MLE of this function of ϑ by substituting $\hat{\vartheta}$ for ϑ. With the data of problem 15, we find the MLE of the theoretical median is $\sqrt{2 \ln 2 (74.51)} = 10.16$, which is actually a very reasonable value for the data since the sample median — the average of the middle order-statistics — is $(10.23 + 10.95)/2 = 10.59$.

(3). (a) The number of the 2000 independently generated intervals which contain the true μ_0 is a $Binom(2000, 0.9)$ random variable, with expected value $2000 \cdot 0.9 = 1800$.

(b). This $N \sim Binom(2000, .1)$.

(c). Probability for a $Binom(2000, .1)$ random variable to fall in $[185, 220]$ is given by the Central Limit Theorem as approximately $\Phi((220 - 2000(.1))/\sqrt{2000(.1)(.9)}) - \Phi((184 - 2000(.1))/\sqrt{2000(.1)(.9)}) = 0.8155$. The continuity-corrected normal approximate value is 0.8128 and the exact binomial probability is

```
pbinom(220,2000,.1)-pbinom(184,2000,.1)
[1] 0.811931
```

(d). This event says that a $Binom(20, .1)$ random variable is 0 (i.e., the number of failures of coverage is 0 out of 20), which is approximately the $Poisson(2)$ probability of 0, or 0.1353. The exact binomial probability of 0 out of 20 is $(0.9)^{20} = 0.1216$.

(4). #8, p. 262. Here the point is that the CI we build — based on specified probabilities α_1, α_2 of μ_0 respectively falling above and below the interval endpoints — is obtained by solving the simultaneous pair of inequalities (which hold with probability $1 - \alpha$)

$$-z_{\alpha_1/2} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha_2/2} \frac{\sigma}{\sqrt{n}}$$

to obtain the asymmetric level-$(1-\alpha)$ CI $(\bar{X} - z_{\alpha_2/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha_1/2} \frac{\sigma}{\sqrt{n}})$. (b). The width of the interval found in (a) is $(z_{\alpha_1/2} + z_{\alpha_2/2})\,\sigma/\sqrt{n}$, while the width of the standard symmetric interval is $2z_{\alpha/2}\,\sigma/\sqrt{n}$. You can see in the specific example $\alpha = .05, \alpha_1 = .0125, \alpha_2 = .0375$ that the symmetric one is shorter, since $2z_{.025} = 3.920$, while $z_{.0125} + z_{.0375} = 1.780 + 2.241 = 4.022$.

(5). #10, p. 262. The idea of this problem is that for Expon(λ) random variables, as in Example 7.5, $2\lambda \sum_{i=1}^{n} X_i \sim \chi^2_{2n} = \text{Gamma}(n, 1/2)$, so that with probability $1-\alpha$, $2\lambda \sum_{i=1}^{n} X_i$ falls between the $\alpha/2$ and $1-\alpha/2$ quantiles of χ^2_{2n}. (a). In this problem, $n = 15$, $\sum_{i=1}^{15} X_i = 63.2$, and the quantiles either from the $\nu = 30$ line of the table on p. 673 or from the **R** values `qchisq(.025,30), qchisq(.975,30)`, are $16.79, 46.98$. Our confidence interval comes by solving the inequalities $16.79 < 2\lambda(63.2) < 46.98$ to obtain the interval $.133 < \lambda < .372$. (b). The only change in this part is that the quantiles become `qchisq(.005,30), qchisq(.995,30)` or $13.79, 53.67$, leading to the 99% interval for λ as $(13.79/(263.2)),\ 53.67/(2(63.2))) = (.109, .425)$. (c) The standard deviation for Expon(λ) lifetimes is $1/\lambda$, and the 95% interval for this parameter is derived from part (a) as $(1/.372,\ 1/.133) = (2.69,\ 7.52)$.

(6). #18, p. 269. Lower (one-sided) 90% large-sample confidence bound is $\bar{x} - z_{0.1}\, s/\sqrt{n} = 4.25 - 1.282(1.30)/\sqrt{78} = 4.061$.

(7). #20, p. 269. This problem calls for a large-sample two-sided 99% confidence interval for a binomial proportion based on $n = 4722$, $\hat{p} = .15$, or $.15 \pm z_{.005}\sqrt{.15(.85)/4722} = (.137,\ .163)$. The fancier CI from formula (7.10) in this problem is $(.137,\ .164)$, hardly any different because n is so large.

(8). R code to input the data, create histograms, overlay normal densities, and provide CI's is:

```
 ## part (a)
> logdat = scan("http://www-users.math.umd.edu/~evs/s401/lgprcp.dat")
> hist(logdat, nclass=20, prob=T, xlab="log precip")  ### I like this one
  curve(dnorm(x, mean(logdat), sd(logdat)), add=T, col="blue")
        ### very non-normal -- asymmetric and multimodal
> mean(logdat)+c(-1,1)*qt(.975,69)*sqrt(0.25/70) ### (b) Normal t-interval
[1] 3.323130 3.561572
> mean(logdat)+c(-1,1)*qnorm(.975)*sqrt(0.25/70) ### (c) Normal z-interval
[1] 3.325221 3.559481
> mean(logdat)+c(-1,1)*qnorm(.975)*sd(logdat)/sqrt(70)
[1] 3.318552 3.566150  ## (d) non-normal unknown; known same as (c)
```