Stat312: Midterm I Solution

Moo K. Chung, Yulin Zhang
mchung@stat.wisc.edu, yulin@stat.wisc.edu

March 5, 2003

1. Let $X_1, \cdots, X_n$ be a random sample from Bernoulli distribution with parameter p.
   (a) What is $\mathbb{E}(S^2/p^2)$? S^2 is the sample variance. Explain your results (10 points).
   (b) Find an unbiased estimator of p^2. Explain your results (5 points).

   Solution. (a) The sample variance is an unbiased estimator of the population variance. Hence $\mathbb{E}(S^2/p^2) = \mathbb{E}(S^2)/p^2 = \mathbf{Var}(X_i)/p^2$. The variance for a Bernoulli random variable can be esily computed as $\mathbf{Var}(X_i) = p(1-p)$. So $\mathbb{E}(S^2/p^2) = (1-p)/p$. (b) We know S^2 and $\bar{X}$ will be unbiased estimators of population variance $p(1-p)$ and mean p respecively. So $\mathbf{E}(S^2) - \mathbf{E}(\bar{X}) = p(1-p) - p = -p^2$. Hence, $\bar{X} - S^2$ is an unbiased estimator of p^2.

2. Let X_1, X_2 be a random sample from $N(0, 1/\theta)$. Note that the sample size is 2 and the density function for X_i is $f(x_i) = \frac{\sqrt{\theta}}{\sqrt{2\pi}} \exp(-\theta x_i^2/2)$.
   (a) Obtain an estimator of θ using the method of moments. Explain your results (5 points).
   (b) Find the likelihood function and use it to obtain the maximum likelihood estimator of θ. (5 points for each question).

   Solution. (a) Match the second moments from the sample and population.

   $$(X_1^2 + X_2^2)/2 = \mathbb{E}(X_i^2) = \mathbf{Var}(X_i) = \frac{1}{\theta}.$$

   Hence $\hat{\theta} = 2/(X_1^2 + X_2^2)$. (b) The likelihood function is $L(\theta) = \theta/(2\pi)\exp(-\theta(x_1^2 + x_2^2)/2)$. Get log-likelihood function $\ln L(\theta) = const + \ln\theta - \theta(x_1^2 + x_2^2)/2$. Differentiate with respect to θ to get the maximum:

   $$\frac{d \ln L(\theta)}{d\theta} = \frac{1}{\theta} - \frac{1}{2}(x_1^2 + x_2^2) = 0.$$

   Solving the equation, we get $\hat{\theta} = 2/(x_1^2 + x_2^2)$.

3. Consider the sample of IQ of 10 dogs: 25, 21, 22, 17, 29, 25, 16, 20, 19, 22. Suppose that these are from a normal population.
   (a) Compute a 95% confidence interval of the mean IQ (5 points).
   (b) Somehow you figured out the variance of IQ to be 9. Compute a 95% confidence interval of the mean IQ. Explain why you have a smaller confidence interval (5 points for each question).

   Solution. (a) Let $X_i \sim N(\mu, \sigma^2)$. Since we do not know σ, we need t-statistic to construct a CI: $t = (\bar{x} - \mu)/(s/\sqrt{n})$. After a lengthy algebra, CI will be $\bar{x} \pm t_{\alpha/2,9} s/\sqrt{10}$. $\bar{x} = 21.6, s = \sqrt{15.6}, t_{\alpha/2,9} = 2.26$. After punching a calculator, CI is $[18.8, 24.4]$. (b) Since σ is known, we need Z-statistic to construc a CI: $z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$. After a lengthy algebra, CI will be $\bar{x} \pm z_{\alpha/2} 3/\sqrt{10}$. CI is $[19.7, 23.5]$. In (a), two parameters are unknown while in (b), only one paramter is unknown. Hence your interval estimation should be more accurate in (b).

4. 47 out of 102 doctors knew the generic name for the drug methadone.
   (a) Compute a 98% confidence interval for the propotion of all doctors who know the generic name for methadone (10 points).
   (b) Compute a 98% confidence interval for the propotion of all doctors who **do not know** the generic name for methadone (5 points).

```
>qnorm(0.02)
[1] -2.053749
>qnorm(0.01)
[1] -2.326348
```

   Solution. (a) Since the sample size is large, CI computation will be based on $Z = (\hat{p} - p)/\sqrt{p(1-p)/n} \sim N(0,1)$. This requires solving a quadratic equation so we will approxmate this by substituting p in the denominator to $\hat{p}$. Then CI becomes $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$. $\hat{p} = 47/102$. So CI is $[0.35, 0.58]$. (b) The propotion of doctors who do know know is $q = 1 - p$. Hence CI would be $[1 - 0.58, 1 - 0.35] = [0.42, 0.65]$.